WorldCat: Estimates of non-acute hospitalization

OCLC 32026949	HEC Holdings - 2 other holdings								
Books	**Rec Stat** c		**Entered** 19950221		**Replaced** 20041231074246.185				
Type a	**ELvl** I		**Srce** d		**Audn**		**Ctrl**		**Lang** eng
BLvl m	**Form**		**Conf** 0		**Biog**		**MRec**		**Ctry** miu
	Cont b		**GPub**		**LitF** 0		**Indx** 0		
Desc a	**Ills** a		**Fest** 0		**DtSt** s		**Dates** 1987,		

040 EYM ‡c EYM ‡d OCLCQ ‡d HEC
043 n-us---
090 RA394 ‡b .S77 1987
090
049 HECA
100 1_ Strumwasser, Ira.
245 10 Estimates of non-acute hospitalization : ‡b a comparative analysis of the appropriateness evaluation protocol and the standardized medreview instrument : final report / ‡c submitted by Ira Strumwasser, Nitin V. Paranjpe and the research staff of the Michigan Health Care Education and Research Foundation, Inc. ... to the Health Care Financing Administration, Office of Research and Development.
260 [Detroit, Mich.?] : ‡b Michigan Health Care Education and Research Foundation, ‡c [1987]
300 3 v. : ‡b ill. ; ‡c 28 cm.
500 Supported by HCFA cooperative grant no. 18-C-98582/5-01 & 02.
500 "September 1, 1987."
504 Includes bibliographical references.
505 0_ Book I -- Book II. Appendix A (Reliability) and appendix B (Validity) -- Book III. Appendix C (Estimates of non-essential acute care hospitlization).
513 Final report; ‡b July 1984-July 1987.
530 Available at cost as a print-on-demand technical [illegible] National Technical Information Service; ‡d PB89-127237 (Book I)
530 Available at cost as a print-on-demand technical rep[illegible] ‡b National Technical Information Service; ‡d PB89-127245 (Book II)
530 Available at cost as a print-on-demand technical report; ‡b National Technical Information Service; ‡d PB89-127252 (Book III)
650 _0 Medical care ‡x Evaluation ‡z Michigan.
650 _0 Hospital utilization ‡x Length of stay ‡z Michigan.
650 _0 Hospitals ‡x Prospective payment ‡z Michigan.
650 _0 Diagnosis related groups ‡z Michigan.
650 _2 Diagnostic Tests, Routine.
650 _2 Length of Stay.
650 _2 Patient Admission.
700 1_ Paranjpa, Nitin V.
710 1_ United States. ‡b Health Care Financing Administration. ‡b Office of Research and Development.
710 2_ Michigan Health Care Education and Research Foundation, Inc.

Action Status	Delete Holdings _	Export **C**	Label _	Produce _	Update Holdings **C**
	Validate _				

MICHIGAN HEALTH CARE EDUCATION AND RESEARCH FOUNDATION, INC.

(313) 225-8706

Mail Code: E-163
600 Lafayette East
Detroit, Michigan 48226

September 10, 1987

James Beebe
Health Care Financing Administration
Room 2-C-16 Oak Meadows Building
6325 Security Blvd.
Baltimore, Maryland 21207

Dear Jim:

Enclosed are three copies of our Final Report on the comparison of two utilization review criteria. I wish to thank you for your concise and helpful feedback on an earlier draft. Please let me know if any questions remain. I believe we answered and addressed each of your comments which, I might add, were most appropriate.

Sincerely,

Ira Strumwasser, Ph.D.
Director of Research

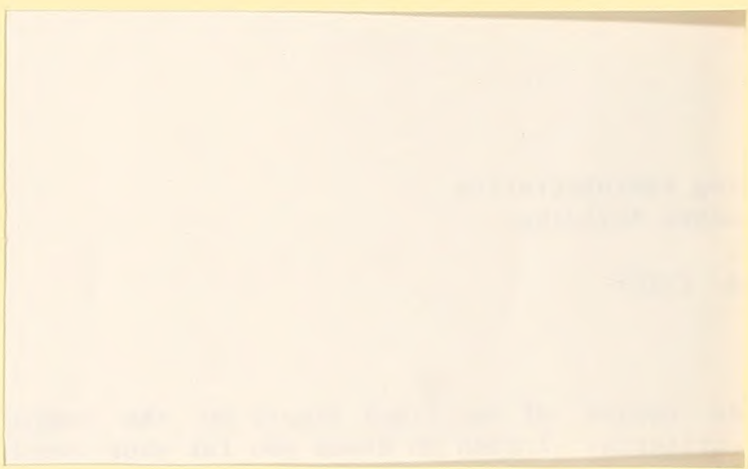
IS:ncs

cc: W. Pembleton (w/attachment)
Sherry Terrell, Ph.D. (w/attachment)

ESTIMATES OF NON-ACUTE HOSPITALIZATION: A COMPARATIVE ANALYSIS OF THE APPROPRIATENESS EVALUATION PROTOCOL AND THE STANDARDIZED MEDREVIEW INSTRUMENT

FINAL REPORT - BOOK I OF III

HCFA Grant No. 18-C-98582/5-01 & 02



ESTIMATES OF NON-ACUTE HOSPITALIZATION: A COMPARATIVE ANALYSIS OF THE APPROPRIATENESS EVALUATION PROTOCOL AND THE STANDARDIZED MEDREVIEW INSTRUMENT

FINAL REPORT - BOOK I OF III

HCFA Grant No. 18-C-98582/5-01 & 02

ESTIMATES OF NON-ACUTE HOSPITALIZATION:

A COMPARATIVE ANALYSIS OF THE APPROPRIATENESS EVALUATION PROTOCOL AND THE STANDARDIZED MEDREVIEW INSTRUMENT

FINAL REPORT

submitted by

Ira Strumwasser, Ph.D.
Director of Research
Principal Investigator

Nitin V. Paranjpe, Ph.D.
Senior Research Analyst
Project Director

and

The Research Staff of the

Michigan Health Care Education and Research Foundation, Inc.
(A supporting, non-profit, 501(c)(3), health services research subsidiary of
Health Service Company
a subsidiary of
Blue Cross and Blue Shield of Michigan)

to the

HEALTH CARE FINANCING ADMINISTRATION
OFFICE OF RESEARCH AND DEVELOPMENT

September 1, 1987

HCFA Cooperative Grant No.: 18-C-98582/5-01 & 02

HCFA Project Officer: James Beebe
Office of Research and Development

TABLE OF CONTENTS

Dedication

For Mr. John C. McCabe, whose vision, and Mr. Richard E. Whitmer, whose leadership, gave life to the Michigan Health Care Education and Research Foundation.

Acknowledgements

The authors wish to express their thanks to the many individuals who gave of their time and talents to ensure a successful project: The following employees of Blue Cross and Blue Shield of Michigan - Willa Ayers, RN; Tom Burzynski, NAAG; Larry Cavallero, Medicare Information Services; Harriett Chenault, RN; Nwamaka Dike, RN; Barbara Dombrowski, RN; Hal Hall, NAAG; Roy Hill, Data Services; Charlotte Holly, RN; Philip Howell, Customer Audit; Ajeet Kang, NAAG; Lynn King, RN; Jolie Laker, RN; Rosalee Livingston, Health Care Affairs; Jack McNally, Customer Audit; Robert Reveley, Vice President, BCBSM, President, Michigan Medical Service (MMS), President, Health Service Company; Melinda Ross, RN; Mary Scheufler, RN; Larry Sell, M.D., Senior Vice President; David Share, M.D., Health Care Affairs; Tom Smith, Provider Organization; Mike Stump, Cost Containment Operations and Charles Zech, Institutional Relations.

Steve Herringa, Ph.D., Institute for Social Research, The University of Michigan; David E. Hetrick, Institute for Labor and Industrial Relations, The University of Michigan; and Graham Kalton, Ph.D., Institute for Social Research, The University of Michigan provided invaluable consultation on various technical aspects of the research.

We also wish to acknowledge organizational support from the Michigan Hospital Association; cooperating hospitals; the Medical Services Administration, Department of Social Services, State of Michigan; and Health Care Affairs (BCBSM), Utilization Review and Quality Control (BCBSM), Customer Audit Department (BCBSM), Cost Containment Program Operations Unit (BCBSM) and National Administrative Auto Group (BCBSM).

The authors also gratefully acknowledge the intellectual, organizational and financial support provided by the MHCERF Board of Directors: Richard E. Whitmer, Chairman, under whose leadership this project was initiated and completed; William S. Hoffman, Ph.D., Secretary; William E. Stevenson, Treasurer; Larry Sell, M.D., Robert Black, M.D., and Franklin McDevitt, D.O.

This research could not have been completed without the commitment to excellence, dedication, enthusiasm and professionalism of three key Foundation personnel. The authors gratefully acknowledge the efforts of Susan Niskanen, Nancy Szydlowski, and Mary Hughes for their tireless assistance over the course of the research.

Finally, the support of Project Officer James Beebe and Sherry A. Terrell, Ph.D. of the Health Care Financing Administration (HCFA) is gratefully acknowledged.

ESTIMATES OF NON-ACUTE HOSPITALIZATION:

A Comparative Analysis of the Appropriateness Evaluation Protocol and the Standardized Medreview Instrument

Executive Summary

The purpose of this research was to assess the reliability and validity of the Appropriateness Evaluation Protocol (AEP) and the Standardized Medreview Instrument (SMI). The utilization review criteria are used by third party payors, health insurance companies and utilization review groups to determine the appropriateness of hospitalization. The criteria were assessed for their accuracy in identifying non-acute medical and surgical admissions and days of stay since it is important that the instruments provide valid and reliable data for audit and decision making purposes. A second focus of the study was to determine rates of non-essential hospital care in southeast Michigan.

Twenty-one hospitals in southeast Michigan were selected for the study. From these, a sample of 1,266 admissions, involving 8,600 days of care, was drawn. Within this sample, all admissions, discharges, and a random selection of days of stay were assessed for the appropriateness of hospitalization. Reliability was established by comparing evaluations by pairs of raters using the same instrument and reviewing the same cases. Validity was assessed by comparing criteria evaluations with the clinical judgment of physicians.

The findings of this research indicate that the AEP is reliable in determining non-acute admissions and days of stay for non-surgical cases. The SMI is unreliable in determining non-acute hospitalization. Preliminary results suggest that the Surgical AEP is less reliable than the AEP and needs refinement before implementation.

Results from the validity trials suggest that the AEP overestimates non-acute care. The SMI was not tested for validity since it was determined to be unreliable. Using AEP findings, adjusted for validity, hospitals in southeast Michigan exhibited rates of non-acute admissions of 12% and approximately 22% of all days of stay.

This research demonstrates that evaluations on the appropriateness of hospitalization can be made by the use of a suitable utilization review instrument, applied by carefully trained health professionals. A reliable instrument like the AEP, even with limits on its accuracy, can be a useful tool in identifying questionable hospitalizations recommended for physician review. The AEP may also be useful in identifying hospitals which exhibit high rates of non-essential hospitalization.

The data presents support for the continuation of predetermination review programs to reduce health care expenditures without reducing the quality of patient care. The study suggests that prospective payment (Diagnostic Related Groups), in conjunction with preauthorization programs, may significantly reduce non-essential hospitalizations and unnecessary health care costs.

ESTIMATES OF NON-ACUTE HOSPITALIZATION:

A Comparative Analysis of the Appropriateness Evaluation Protocol and Standardized Medreview Instrument

Ira Strumwasser, Ph.D.
Director of Research

Nitin V. Paranjpe, Ph.D.
Senior Research Analyst

and the Staff of the

Michigan Health Care Education and Research Foundation

A supporting, non-profit, 501(c)(3) health services research and philanthropic subsidiary of Health Service Company

(A subsidiary of Blue Cross and Blue Shield of Michigan)

ABSTRACT

The purpose of this research is to assess the reliability and validity of the Appropriateness Evaluation Protocol (AEP) and the Standardized Medreview Instrument (SMI) in identifying non-acute admissions to, and days of care in, acute care hospital settings. A second focus is to estimate rates of non-acute hospital care in southeast Michigan.

A probability sample of 1,266 discharges from hospitals in southeast Michigan was drawn. In the sample selection, hospitals were stratified on: (1) bed size, (2) teaching status and (3) occupancy rate. Twenty-one hospitals were included in the sample. A random selection of days of stay (including all admission days and days of stay before the actual discharge day) was drawn from the sampled records.

Analysis focused on the reliability and validity of each criteria-set and an analysis of the rates of non-acute care. Physician reviews of a subset of records served as validation for determining the accuracy of the criteria.

Reliability Results:

The Standardized Medreview Instrument, when applied by RNs newly and extensively trained in applying the criteria, is unreliable in assessing both the necessity of admissions and the need for continued days of stay. The data

also indicate that extensively trained RNs, working together in one organizational setting, with at least two years of experience with the SMI, are able to apply the SMI criteria with more consistency. However, this consistency is below that obtained when the Appropriateness Evaluation Protocol is applied by RNs newly and extensively trained in its application. The AEP is a reliable instrument for identifying non-acute medical admissions and days of care in acute care hospitals. The Surgical AEP (SAEP) is less reliable in assessing surgical admissions and days of stay within surgical admissions.

Validity Results:

Since the SMI was not reliable, it could not be valid. Thus, assessment of the accuracy of the criteria for identifying non-essential hospitalization focuses on the AEP/SAEP criteria. The validity or accuracy of the AEP criteria in identifying non-acute hospital care was as high as 71% for both admission and day review. When physicians disagreed with the findings of the nurses applying the AEP criteria, the disagreements were almost always false positives. Physicians determined care to be acute while the AEP criteria identified the care as non-acute. The AEP was found to overestimate unnecessary hospitalization by approximately 40%.

Implications:

The findings are important when considering the use of these utilization review criteria by third party payors (Blue Cross), insurance companies and utilization review groups, such as professional review organizations (PROs), to identify non-acute care and to determine the appropriateness of insurance payments and hospitalization in managed care programs (i.e., preadmission review). The tendency toward overestimation of non-acute care (false positives) suggests important differences in application in these utilization review contexts.

First, in utilization review studies designed to assess the degree of non-acute hospitalization, it should be noted that the rates of non-acute care produced by a straightforward application of the AEP will overestimate unnecessary hospitalization. Thus, these rates should be reduced by an appropriate validation factor. Second, in pre-determination programs, it is vital that non-acute cases be referred for physician review. **To maintain cost effective, quality care, it is essential that 100% of the denials for preadmission certification be reviewed by expert utilization review physicians.**

The quality of utilization review and cost containment efforts depends on the review instruments selected by insurance, third party payor, or utilization review organizations. Quality utilization review programs depend on physician involvement for second opinions, when the first opinion is based on instrument assessment. Safeguards, such as appeals processes, are vital to ensure quality and cost effective utilization review and cost containment programs.

Estimates of Non-Essential Hospitalization:

The data show that physician adjusted rates of non-acute care were approximately 12% of the admissions and 22% of all days of stay in 1983. Rates of non-acute care within medical admissions were higher than non-acute care within surgical admissions. Forty-three percent of all non-acute days were found in non-acute admissions. In general, teaching facilities appear to have lower rates of non-acute care than non-teaching facilities. These data argue for the continuation of utilization review and cost containment programs and underscore the need for predetermination reviews. DRG payment, in conjunction with preauthorization programs, may have a significant influence in reducing non-acute hospital admissions and unnecessary health care costs without adversely affecting the quality of health care.

Future Research:

These data are based on 1983 rates of non-acute care. Since a host of utilization review and cost containment activities in the private sector and the federal government have been initiated, e.g., prospective payment and preadmission review programs, additional research should be conducted to determine the rates of non-acute care in 1986. Finally, a third utilization review criteria set, the Intensity, Severity, Discharge (ISD) Criteria, developed by InterQual should be assessed for its reliability and validity.

CHAPTER 1

INTRODUCTION

The high costs of health care are well documented. In 1982, national health care expenditures reached a record 10.5% of the Gross National Product (GNP). Hospital care expenditures accounted for approximately 42% of total health care expenditures (Gibson, Waldo & Levit, 1983). Approximately 40% of total inpatient expenditures was derived from government health programs (Medicare and Medicaid). Estimated 1982 calendar year national expenditures for inpatient care under the Medicare program was 34 billion dollars while inpatient hospital care costs for the Medicaid program amounted to roughly 8 billion dollars. Given such high costs, the notion that some of these expenditures might be unnecessary deserves considerable attention.

As resources become more and more limited, controlling health care costs remains an ever present dilemma. A constructive means of reducing costs while maintaining or enhancing the quality of health care, is to target non-acute hospital use, which ensures that cost savings are not achieved by limiting

access to necessary acute care. Concern by business, labor, third party payors (BCBS plans), the federal government, hospital administrators, consumers and physicians over the high cost of health care makes the development of reliable and valid criteria designed to assess the appropriateness of the setting in which acute care is provided of prime importance.

Published evidence suggests that non-acute[1] hospital use may be a significant phenomenon. Table I - 1 presents a summary of the major reported studies of non-acute hospital use in the United States. These studies indicate 6% to 20% of hospital bed utilization to be non-acute. While these studies have employed diverse methods ranging from level of care and specific medical criteria to the use of individual observations and group Delphi techniques, their similar findings of perceived non-acute use is striking despite their methodological differences.

Insert Table I - 1 About Here

---

[1] In this context, the term non-acute refers to hospital care that could have been provided in other than an acute care hospital setting. Throughout this report, the term "non-acute hospital care" is intended to refer to the location of, rather than the quality or appropriateness of, services. The terms unnecessary or inappropriate hospital care, frequently used to describe non-acute care, are misleading to the extent they suggest that care was of a poor quality or that the care itself was not required. The utilization review criteria employed in these studies assess the appropriateness of the setting or location of services and not the appropriateness or quality of the care itself.

TABLE I - 1

CASE STUDIES OF NON-ACUTE HOSPITAL BED USE*

Study Author(s)	Hospital Study Population	% Non-Acute Use
Querido (1963)	General Hospital Services, 20 Amsterdam Hospitals	17%
Browning & Crump (1969)	Medical/Surgical Services, Rochester, N.Y., Hospitals	14%
Gertman & Bucher (1969)	Medical Services, Baltimore City Hospital, Baltimore, Maryland**	12%
Zimmer (1974)	All Clinical Services, Strong Memorial Hospital, Rochester, New York***	9%
	(Medical Service Alone)	(12%)
Restuccia & Holloway (1976)	Medical/Surgical Services, Herrick Memorial Hospital Berkeley, California****	11%
Restuccia et al. (1984)	National Sample	19.1% (admissions)
		20% (of days in acute admission)
SysteMetrics (Moynihan et al., 1984)	National Sample	5.7% (admissions)
		7.8% (of days in acute admission)

* Adapted (and modified) from Table 1, pg.5, Gertman, P.M. and Restuccia, J.D. "Methods to Determine Inappropriate Use of Hospital Services: The Appropriateness Evaluation Protocol (AEP)", Final Report submitted to the HCFA: Grant No.: 18-P-97513/1-02.

** Major medical school hospital.

*** Community hospital.

**** Based on total available bed days, not appropriateness of actually utilized days; correction would make result = 20%.

The principal methodological concern in case studies on the use of hospital beds, during the last twenty years, has been reviewer reliability (see Methods section for definition of reliability). As shown in Table I - 2, most methodological studies, with the exception of Gertman and Restuccia (1981) and SysteMetrics (1984), have exhibited difficulties in obtaining reliable measures of the need for hospitalization. While reviewers tended to have moderately high levels of overall agreement, the methods were unreliable.

Insert Table I - 2 About Here

The reason for the lack of reliability is that a large proportion of overall agreement was due to chance because most days were judged to be acute by both reviewers. Table I - 2 illustrates that, with the exception of the Moorehead; Gertman and Restuccia; and SysteMetrics studies, no unstructured approach (i.e., using reviewers' subjective judgment) has achieved a 50% level of agreement between two or more judges on which days are considered to be non-acute.

TABLE I - 2

EXAMPLES OF STUDIES ON RELIABILITY PROBLEMS IN JUDGING

NON-ACUTE USE WITH RELATIVELY UNSTRUCTURED SUBJECTIVE CRITERIA*

Author(s) and Nature of Study	Percent of Cases Questioned by One or More Judges	Agreement as Percent of All Cases (Overall Agreement)	Agreement as Percent of Questioned Cases (Specific Agreement)
Moorehead (1964): two judges	26	88	53
Browning (1965): two judges	10	86	20
Gertman and Restuccia (1981): two judges	--	91.8 - 94.3	58 - 72
SystoMetrics (1984): two judges	--	98.7	62.5
Zimmer (1967): charts only, two judges	19	85	23
Zimmer (1967): pooled data, four judges	27	73	6
Zimmer and Groomes (1969)**: two outside physicians	8	84	0
Zimmer and Groomes (1969)**: two nurses examining same cases as above	20	76	40

* Adapted and modified from Table IV.62, page 386, Donabedian, A., Aspects of Medical Care Administration, Harvard University Press, Cambridge, Massachusetts, 1973.

** Chronic Care Facilities.

The Appropriateness of Acute Care Hospitalization: "State-of-the-Art" Criteria-Sets

Two instruments for assessing non-acute hospitalization, the Appropriateness Evaluation Protocol (AEP) and the Standardized Medreview Instrument (SMI) developed by Gertman and Restuccia (1981) and SysteMetrics (1984), respectively, have recently emerged as reliable (defined as the extent to which the assessment of one reviewer is replicated by another reviewer or the consistency of the criteria's application) and valid (defined as the agreement between reviewers using the criteria and physician assessments of the necessity of hospitalization) for measuring the appropriateness of the setting in which acute care is provided.[2] However, these instruments report markedly different rates of non-acute admissions and days of stay (DOS) for adults receiving medical, surgical and gynecological medical services in acute care, short-stay hospitals. Application of the AEP indicates that, in general, the range of non-acute admissions is between 10 and 35 percent (depending on the particular study); while non-acute days of care represent approximately 20-30 percent of all days of care within otherwise acute admissions. The SMI finds the rate of non-acute admissions to be 5.7 percent and the rate of non-acute days of care (DOC), within otherwise acute admissions, to be approximately 7.8 percent. Given the importance of developing reliable and valid (see validation results for definition) measures of the efficient use of hospital beds, and the differences in the reported magnitude of non-acute care found by

---

[2] Strumwasser, I. and Paranjpe, N.V. (in progress) are conducting an evaluation of Interqual's Intensity, Severity, Discharge utilization review criteria-set.

application of these instruments, it is critical to determine which, if either of the two, is the most reliable and valid. The following section describes each instrument, major findings and the critical substantive and methodological issues regarding the AEP and SMI.

The Appropriateness Evaluation Protocol (AEP)

The developers of the AEP claim that it is a diagnostic-independent, objective, criteria-based technique for determining the medical necessity of hospital admissions and days of care for all adult medical, surgical and gynecological patients. The AEP is a systematic method of aiding reviewers (usually RNs) in determining the reasons for non-acute admissions and days of care. Research employing the AEP indicates that non-acute admissions range from 10 to 35 percent and rates of non-acute days of hospital care, within both acute and non-acute admissions, range from 10 to 50 percent (averaging approximately 25% of all DOC). While the rates of non-essential hospital care reported with the use of the AEP vary depending on the specific study cited, in general most studies report rates of 10 to 15 percent of hospital admissions and 20 to 30 percent of all days of care (within otherwise acute admissions) to be non-acute. (See Blumenfeld, 1983 for a critique of the AEP Technical Report.)

The AEP is a list of 27 criteria for DOS (see Attachment A) that describes medical services, nursing services, and patient conditions such that meeting any one of the criteria represents sufficient reason to continue treatment on an inpatient basis. Any patient who, on a given day, does not meet any of these criteria would then be defined as receiving non-acute hospital level

care. The first two criteria batteries (medical and nursing) (see Attachment A, AEP, part VII, A, Medical Services and B, Nursing/Life Support Services) consist of services that are typically provided only at an acute level of hospital care. The third battery (Attachment A, AEP, part VII, C, Patient Condition) includes factors that indicate immediate acute care hospitalization is necessary based on patient condition. There are 18 admission criteria organized such that meeting any one of the severity of illness or intensity of service criteria justifies admission to an acute care facility.

The Standardized Medreview Instrument

The Standardized Medreview Instrument (SMI) developed (under contract with HCFA) by SysteMetrics, Inc. (SysteMetrics, 1984) is, according to SysteMetrics (1984), a criterion-based, medical records review instrument developed to measure the appropriateness of admissions and days of care. The instrument contains 117 admission criteria and 56 criteria for days of care (30 items for acute care service and 26 items related to patient condition) that indicate the need for continued hospitalization (see Attachment B). The SMI, like the AEP, was designed to be used retrospectively with medical records. However, the instrument developers claim that both instruments may also be used with medical records on a concurrent, or pre-admission, pre-authorization utilization review basis. (See Blumenfeld, 1985 for a critique of the SMI Technical Report.)

The developers of the SMI operationally define unnecessary utilization to occur when: (1) a patient is hospitalized unnecessarily (as in an admission for a procedure which could be performed on an outpatient basis or for

treatment which could be rendered in a sub-acute nursing facility or in a physician's office), (2) the patient is kept in the hospital after he/she is ready for discharge or transfer to a lower level of care, or (3) services are delivered inefficiently, so that the hospital stay is unnecessarily prolonged (e.g., when the patient is unnecessarily kept in the hospital several days before a surgical procedure is performed). When using the SMI criteria-set, for hospitalization to be acute, both patient condition factors and level of service factors should be met, i.e., the patient should be receiving acute level services and should actually have a condition (whether longstanding or in an acute or exacerbated phase) which requires those services. SysteMetrics, employing the SMI, found significant but not dramatic rates of non-acute hospital utilization amounting to approximately 5.7 percent of all admissions and 12 percent of all days of care (within both acute and non-acute admissions).

Non-acute Hospital Care: The AEP and SMI

A main focus of this research is measurement of the instruments' specific non-acute reliability. Specific non-acute reliability is the degree of agreement between raters on all cases in which either rater assessed the hospitalization to be non-essential (the number of cases both raters find non-acute divided by the total number of cases either rater finds non-acute).

Although both the AEP and SMI report overall inter-rater reliability coefficients in excess of 80% and specific reliability in excess of 60%, differences in the rates of non-acute hospitalization between the two instruments are substantial. As indicated in the margins of Table I - 3, the AEP finds, on average (an average estimate based on the rates reported in published studies

using the AEP), 12.5 percent non-acute admissions (Gertman and Restuccia, 1981); while application of the SMI to a national sample finds 5.7 percent of all admissions to be non-acute. (It should be made clear that the admission and DOS cell rates in Tables I - 3 and I - 4 are hypothetical based on actual marginal rates found in studies by instrument developers.) [The problem in reporting one rate of non-essential hospitalization using the AEP is that individual studies report different rates of non-acute care. We have chosen what we consider to be generally representative AEP rates of non-acute admissions and DOS. However, even our own internal references to rates will vary, from time to time, as we focus on different studies of non-acute care using the AEP.]

Insert Table I - 3 About Here

As indicated in Table I - 4, on average, the AEP finds that 25 percent of hospital days are non-acute; while the SMI finds that 12 percent of all days of care are non-acute.

Insert Table I - 4 About Here

TABLE I - 3

Non-Acute Admissions:
HYPOTHETICAL* Cell Rate Differences

		Admissions AEP: Acute Admission	Admissions AEP: Non-Acute Admission	Total
SMI	Acute Admission	87.5%	6.8%	94.3%
	Non-Acute Admission	0%	5.7%	5.7%
	Total	87.5%	12.5%	100%

* Based on published papers and/or technical reports produced by the instrument developers.

TABLE I - 4

Days of Care:
HYPOTHETICAL* Cell Rate Differences

Days of Care

AEP

SMI		Acute Days	Non-Acute Days	Total
	Acute Days	75%	13%	88%
	Non-Acute Days	0%	12%	12%
	Total	75%	25%	100%

* Based on published papers and/or technical reports produced by the instrument developers.

Given the importance of reducing non-acute hospital care in terms of both the cost and quality of care, differences in rates of non-acute care reported by the AEP and SMI are critical. If the "real" rate of non-acute admissions is 12.5 percent, then a considerable reduction in non-acute care may be possible via careful medical management and close utilization review. If, on the other hand, the "true" rate of non-acute admissions is 5.7 percent, then a significant reduction in non-acute admissions may be difficult or even undesirable. Moreover, the difference in rates of non-acute hospitalization (6.8%) between the AEP and SMI is troublesome. Employment of the AEP (if the SMI is correct) will result in incorrectly identifying acute admissions as non-acute (a false-positive finding). Similarly, use of the SMI (if the AEP is accurate) will result in incorrectly identifying non-acute admissions as acute (a false-negative finding). These findings have very different implications for interventions employed as a result of utilization review practices. Inappropriate interventions may be unjust and may have a profound negative effect on the quality of health care. Knowledge of the direction of the errors is important for those involved in cost containment activities and for physicians and hospitals that may be asked to monitor their decision-making processes regarding admissions and discharges.[3]

---

[3] This may be especially important under the new Diagnostic Related Groups (DRGs) prospective reimbursement system, especially for Medicare and Medicaid payments, as hospital administrators and PROs begin to monitor physician discharge patterns more closely.

The differences in rates of non-acute days (17.2%) are even more dramatic than the differences in admissions. If 1982 HCFA hospital payments (Medicare and Medicaid) were reduced 17 percent, it would result in potential savings to the federal government amounting to roughly 7 billion dollars annually. Given these reported differences in the magnitude of non-acute care, we now turn to some possible explanations of the differences in the rates of non-acute care between the AEP and SMI.

Possible Reasons for Differences in Non-acute Care Found by the AEP versus SMI

The first, and intuitively obvious reason for differences in rates of non-acute care is that the instruments differ in their respective validity. It has been suggested that the AEP overestimates the magnitude of non-acute care by employing inappropriately stringent criteria. Others have argued that the SMI underestimates the extent of non-acute care by including a host of excessively lenient criteria. To test the validity of the AEP and the SMI it is vital to compare the findings of the criteria-sets with independent physician reviews.

Second, differences in non-acute admissions and days of stay between the AEP and SMI may be due to differences in the populations examined rather than differences in the instruments' criteria. The SMI was applied to a random national sample of hospitals stratified on: region, bed size, ownership, occupancy rate, teaching status and rural/urban differences. In general, the AEP has been standardized on and applied to teaching hospitals. In addition, the AEP has been applied primarily to focused hospital reviews in areas suspect of high rates of non-acute care. [Although a recent report by

Restuccia et al. (1984) presents evidence, based on a random national (geographic) sample, that approximately 20% of all days within otherwise appropriate admissions appear to be non-acute.] Thus, differences in rates of non-acute care may, in part, be due to systematic differences in the populations examined. Application of both the SMI and AEP to a randomly selected sample of hospitals, discharges within hospitals and days of care within discharges will resolve whether population differences account for the differences in non-acute hospitalization reported in research using these two criteria-sets. Holding population constant will allow observed differences in rates of non-acute care to be attributed to the validity of the instruments and/or (rater) reliability rather than population differences.

The third possible reason for differences between the AEP and SMI is rater-reliability. Several possibilities exist. First, it is possible that one or both instruments are unreliable. The critical measure of reliability is specific non-acute reliability, or the ability of raters using the criteria-sets to agree on specific instances of non-acute care. It is possible that, in instances where the AEP and SMI disagree, the inter-rater reliability coefficient may also be low. The differential reliability of the instruments may depend on the difficulty of the case assessed. One or both criteria-sets may have extremely reliable findings in cases clearly acute (e.g., admission for a myocardial infarction or for ketoacidosis and diabetic coma) or those clearly non-acute, (e.g., admission for low back pain for bedrest and empirin with codeine #3 QID with nursing notes that patient is ambulating frequently). One or both instruments may be unreliable in the more difficult cases in which the severity of illness and need for intensive services at the acute-care level

are unclear or highly discretionary (depending on practice patterns and community standards). To test for these possibilities, a sample of cases in which both instruments agree that care is acute and both instruments agree that care is non-acute, as well as instances in which the criteria-sets disagree on the appropriateness of hospitalization, should be subject to re-analysis by a second independent review. If specific reliability is high, then any differences between the criteria may be attributed to instrument validity rather than rater reliability.

An independent assessment of the reliability and validity of the AEP and SMI is essential. If physicians are being asked to modify their practice patterns or if insurance payments are based on the findings of such criteria, it is critical that such decisions be based on reliable and valid data, or the quality of health care may be affected in adverse ways. If costs are to be contained, methods for containing costs while maintaining quality of care are essential. Finally, to measure the effect of the Prospective Payment System (PPS) on admissions to and continued DOS in acute-care facilities[4], reliable and valid measurements must be employed. After presenting comparative data on the relative reliability and validity of the AEP and SMI, we shall present our findings on the estimates of the level of non-acute hospital care in a major six (6) county urban area accounting for roughly 45% of all acute care admissions to all acute care hospitals in the State of Michigan.

---

[4] Strumwasser, I. and Paranjpe, N.V. "Non-acute Hospitalization: Then (1983) and Now (1986)," in progress.

CHAPTER II

METHODS

METHODS

Summary of Methods

The design originally sampled 80 out of a universe of 183 hospitals in the lower peninsula of Michigan. However, financial constraints led to a subsample of 40 out of 80 hospitals in By-Law District II (a six county area in southeast Michigan). District II contained 689,564 admissions which was 53.28 percent of the total number of (1983) admissions to all acute-care, non-specialty hospitals in the lower peninsula of Michigan.

Hospitals were stratified by location (geographic area), bed size and occupancy rate using the Controlled Selection Process (Hess, 1975). All hospitals in the area under study with total annual admissions in excess of 13,000 were included in the sample with certainty (self-representing). Non-self-representing hospitals were sampled using the Controlled Selection Process with probability of selection proportional to size of hospital (i.e., large bedded hospitals with more annual admissions had a greater probability of being selected and included in the sample). Records were selected to yield

a self-weighting sample. The average number of records reviewed in self-represented and non-self-represented hospitals were 31.44 and 31.07, respectively.

All admissions and discharge days (the day before actual discharge) were included in the DOS sample with certainty. A maximum of seven days of stay (including admission and discharge) were sampled. In all stays seven days or less, all DOS were reviewed. In LOS greater than seven days, the days between admission and discharge were randomly sampled (maximum of five randomly sampled days). The hospital participation response rate was 21 out of 40 (52.5%), which, as described below, is generalizable to the six county area in southeast Michigan.

Eight RN reviewers were selected and trained in the use of the criteria (four RNs were trained only on the application of the AEP and four RNs were trained on the use of the SMI). A two day training course was conducted by representatives of the instrument developers (for the AEP the trainers were Joseph Restuccia, Dr.P.H. and Bernard Kreger, M.D.; while the trainers for the SMI were Kathy Barnes, ART and Pat Loch, RN of SysteMetrics). RN reviewers applied only the instrument (either AEP or SMI) for which they received training. The RN reviewers were randomly assigned to either the AEP or SMI condition. Quality and not quantity of reviews was stressed. RNs were paid on an hourly or salaried basis and allowed to work at their own pace. Re-reviews (inter- and intra-rater reliability) were conducted by members of the same team. Cases were assigned to the same team member (intra-rater reliability) or different team members (inter-rater reliability).

This chapter presents a description of the methodology used to select the hospital and admission sample tested in this research project. Described are the approaches to sampling hospitals from all hospitals within the state of Michigan, admissions within hospitals, and days of stay within each admission. A description of the hospital sample by various hospital characteristics is also provided. The sample of hospitals suggested by theoretical considerations (e.g., size of hospital, occupancy rate, etc.) is compared with the actual hospital frame from which the sample was drawn. Finally, the set of 21 hospitals included in the study is presented. Because of inaccurate estimates of the time needed to conduct chart reviews, the study proceeded on a reduced sample (i.e., number of reviews performed). Initial resource allocation was based on reviews requiring about 15 minutes per chart. Actual reviews took an average of 40 minutes per chart, which included an admission review and the review of up to 6 days of stay (including the day before discharge).

In addition, since participation in this study was voluntary, non-response (refusal to participate in the study) from hospital administrators resulted in divergence from theoretical to actual sampling distributions. Given these limitations, the sample that evolved was selected for its geographic and economic importance (the high percentage of admissions in the region in comparison to statewide admissions). The area sampled is the southeastern part of Michigan, By-Law District II (see Map II - 1).

Insert Map II - 1 About Here

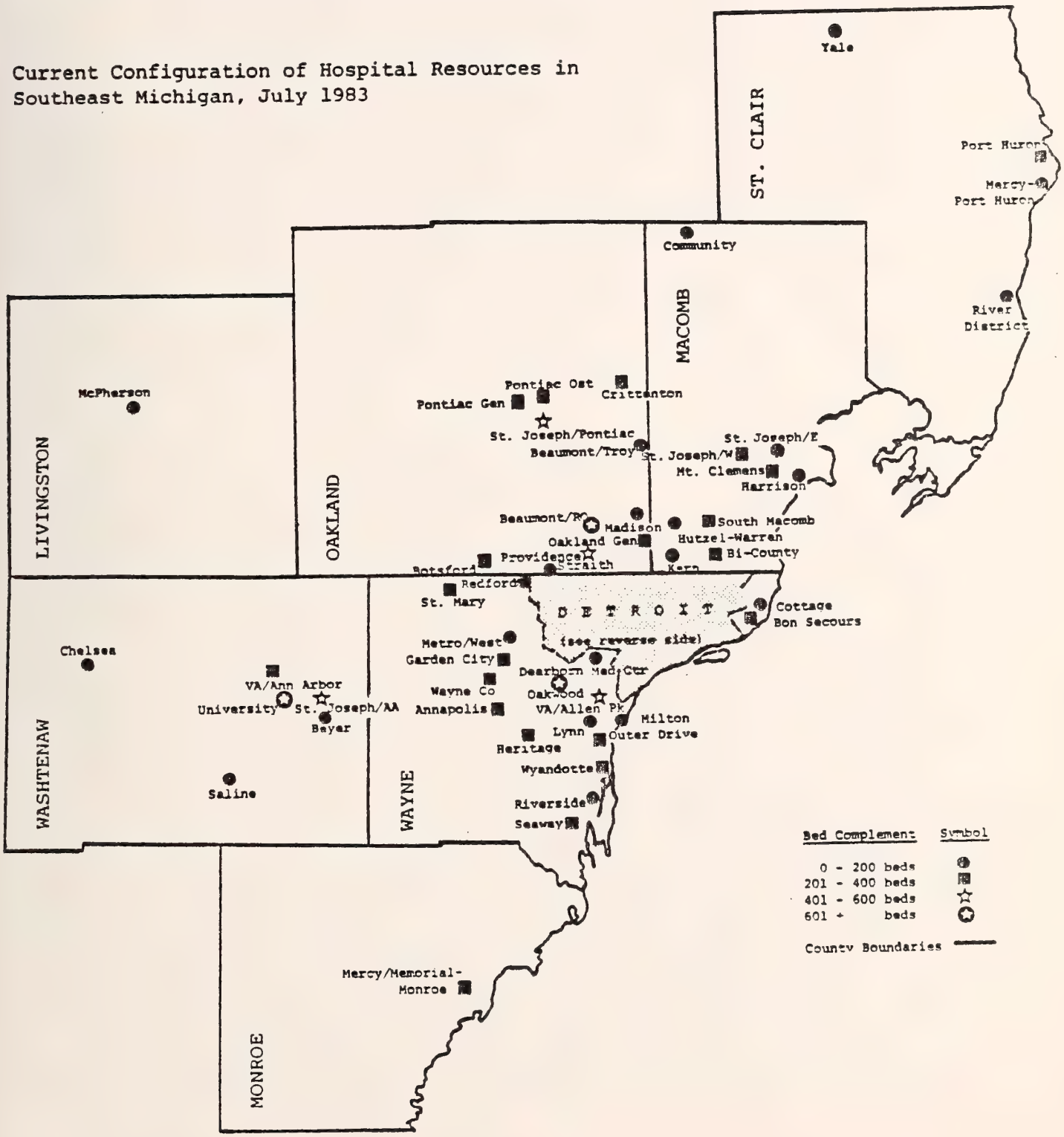

Current Configuration of Hospital Resources in Southeast Michigan, July 1983
ST. CLAIR
Yale
Port Huron
Mercy-Mt. Clemens
Port Huron
MACOMB
Community
River District
LIVINGSTON
McPherson
OAKLAND
Pontiac Ost
Crittenton
Pontiac Gen
St. Joseph/Pontiac
Beaumont/Troy
St. Joseph/E
St. Joseph/W
Mt. Clemens
Harrison
Beaumont/R
Madison
South Macomb
Oakland Gen
Hutzel-Warren
Providence
Bi-County
Kern
Botsford
Redford
St. Mary
DETROIT
Cottage
Bon Secours
(see reverse side)
Chelsea
Metro/West
Garden City
Dearborn Med Ctr
VA/Ann Arbor
University
St. Joseph/AA
Bayer
Wayne Co
Annapolis
Oakwood
VA/Allen Pk
Milton
Lynn
Outer Drive
Heritage
WASHTENAW
Saline
WAYNE
Wyandotte
Riverside
Seaway
Bed Complement
Symbol
0 - 200 beds
201 - 400 beds
401 - 600 beds
601 + beds
County Boundaries
Mercy/Memorial-Monroe
MONROE

Chart reviews were conducted in each of the sample hospitals. The AEP and SMI were applied to the same set of randomly drawn charts based on hospital admission size, and divided proportionately across three payor groups: Medicare, Medicaid, and Blue Cross. A sub-sample of charts was drawn from this population based on disagreements between the criteria on the appropriateness of the admission. (The actual selection procedure and sample description may be found both here and in the reliability and validity results sections.) This sample of charts was drawn with the objective of subjecting the instruments to stringent reliability and validity testing. Finally, since the rates of non-acute care differed markedly between the two criteria, it was necessary to over-sample cases found to be non-acute.

Reliability Methods

To assess the reliability of the AEP and the SMI, 99 charts were selected from the first 13 hospitals studied. An average of approximately eight (8) records were selected for reliability review from each of these 13 hospitals. This constitutes a 7.82% sample of all cases reviewed over the 21 hospitals included in the study. Since these records would be used for both the reliability trials and subsequent validity assessment, they were purposely chosen to reflect instances in which the AEP and the SMI: (1) agreed that an admission was acute, (2) agreed that an admission was non-acute, and, (3) instances in which the two instruments disagreed on the appropriateness of the hospitalization. (Please note the reader is referred to Chapter III for Tables III - 5 through III - 8.) Table III - 5 indicates the distribution of cases on this admission agreement/disagreement model (n = 99) selected for the reliability reviews while Tables III - 7 and III - 8 present the total number

---

Insert Tables III - 5, III - 7 and III - 8 About Here

---

of randomly selected cases reviewed for the estimates of non-acute care (N = 1,173). Since one of the primary reasons for conducting the study was to assess the validity of each criteria set, especially on non-acute findings, the largest number of cases selected for reliability trials and subsequent validity assessments were from the cells in which the two instruments disagreed. The most frequent direction for disagreement between the two criteria-sets was the instance in which the AEP found the admission to be non-acute while the SMI found the admission to be acute. The majority (68.69%) of sampled cases chosen for reliability review were selected from this intersection (see Table III - 5). Because of this sampling procedure, we will provide both unweighted and subsequently weighted reliability coefficients. (See Technical Note on weighted reliability, Attachment D). Since it was also important to determine whether or not instruments were reliable and valid in instances in which they both agreed that care was either acute or non-acute, a small sample of charts was selected from these cells. In general, the selection of cases for duplication and subsequent re-review was based on the following rules: Of all records reviewed within a given hospital, one case was randomly selected in which both instruments agreed the admission was acute. Of all records reviewed within a given hospital, two cases were randomly selected in which both instruments agreed the admission was non-acute. One case, should it be found to exist, was randomly selected in which the SMI found the admission to be non-acute while the AEP found the admission acute (this infrequently occurred since the SMI is more likely to find care to

be acute than the AEP). Finally, at least five cases in which the two instruments disagreed on the appropriateness of the admission and where the AEP found the admission non-acute while the SMI found the admission acute were randomly selected for inclusion in the reliability sample. This sample constitutes the subset of 99 cases (Table III - 5) selected for the reliability trials.

It should be noted that the reliability subset sample is a purposefully biased sample. The reliability subset is biased toward cases in which the admission is found to be non-acute by one (the AEP) or both instruments. For example, 79.79% of the admissions were non-acute according to the AEP while 14.14% of the admissions were non-acute according to the SMI. Thus, the unweighted reliability coefficients are likely to reflect extremely low agreement for instruments that do poorly on non-acute assessments. Since the primary purpose of these criteria-sets is to provide reliable and valid assessments of non-acute care, this biased subset seems appropriate. It is also possible that cases in which the admission is considered to be non-acute by the AEP and acute by the SMI are among the more difficult cases to evaluate thereby testing the performance of the protocols under the most rigorous conditions. However, for population estimates, reliability will vary as a function of the extent of acute and non-acute care reported by each criteria-set. Thus, to approximate expected population reliability coefficients, weighted reliability coefficients will also be provided as a technical note (see Attachment D).

The initial review of charts was performed on-site in each of the 21 hospitals participating in the study. In total, 1,266 unique in-hospital chart reviews using each instrument were performed. (However, both instruments were applied

to only 1,173 identical cases.) These reviews took place between February 5, 1985 and May 31, 1985. The 99 charts selected for inter-rater and intra-rater reliability reviews were from the first 13 hospitals in which charts were reviewed. All reviews for the 99 charts chosen for the reliability assessment took place during February, March and April, 1985.

Two forms of rater reliability were examined. Inter-rater reliability reflects the consistency of reviews between two (2) different raters. Intra-rater reliability reflects the consistency of reviews within each rater. Table III-6 illustrates a hypothetical relationship and interpretation of the interaction between the inter- and intra-rater reliability.

Insert Table III - 6 About Here

Inter-rater reliability assesses the reliability of reviews between raters who may have different levels of understanding, training, and ability, and between reviewers who may interpret the application of the instruments differently. Intra-rater reliability holds individual rater differences constant. It assesses the degree of replicability while holding idiosyncratic rater differences constant. As suggested in Table III - 6, for the application of an instrument to be considered highly reliable, both inter- and intra-rater reliability should be high. Unreliable application of an instrument is reflected by both low inter- and low intra-rater reliability. Contrasting inter- and intra-rater reliability provides information on the source or cause of low reliability, e.g., internal consistency/inconsistency (intra-rater reliability) versus external consistency/inconsistency (inter-rater reliability).

Validity Methods

Each of the cases sampled for reliability and validity (N = 172), described previously, were reviewed by twenty two (22) physicians. Eleven physicians (seven internists and four surgeons) were selected from a Michigan based, staff model, predominantly fee-for-service, large bedded, high quality teaching hospital located in the same community from which the sample cases and hospitals were selected (Henry Ford Hospital, Detroit, Michigan)* and eleven physicians (seven internists and four surgeons) selected from a west coast, high quality, well respected, group model, health maintenance organization (Kaiser Permanente Medical Group, Oakland, California).** The Michigan based physicians' assessments reflected general community medical standards, while the HMO based physicians assessed the appropriateness of care based on HMO practice and program standards. (Kaiser Permanente Medical Group of Northern California was chosen because they have demonstrated, in the published research literature, the lowest rates of hospitalization per 1,000 members of any HMO in the nation.) While it could be argued that it is inappropriate for California based physicians to evaluate care performed in Michigan settings, the idea was to: (1) Assess whether any differences exist within and between Michigan based FFS physicians and California based HMO physicians, and (2) Determine the upper (HMO physicians) and lower limits (FFS physicians) of validity and estimates of non-acute care.

---

* Special thanks to Wilmer Rudd, M.D. and James Bridges, M.D. for their assistance.

** Special thanks to Bruce Sams, M.D. and Richmond Prescott, M.D. for their cooperation and hospitality.

Each of the 172 cases sampled for validity analysis were reviewed by three FFS and three HMO physicians. Internists reviewed patient cases admitted for medical reasons, while surgeons reviewed cases admitted for elective surgery or cases in which a major surgical procedure was performed. Medical admissions for invasive diagnostic or treatment procedures (e.g., bronchoscopy) or for medical admission workup that resulted in surgery were also reviewed, where appropriate, by the surgeons. No attempt was made to match physician specialty with the diagnosis of the cases they reviewed. All physician assessments were independent and the cases assigned to them were random. Physicians were instructed to evaluate the appropriateness of each admission and assigned days of care review based on their independent clinical review. All physicians conducted blind reviews (i.e., physicians did not know whether the care had been deemed to be acute or non-acute by the instruments).

Confirmation of Acute/Non-Acute Care

Confirmation or validation of the criteria is not a straightforward matter. Physicians do not necessarily agree on whether an instance of care is acute or not. Since the leniency or stringency of the validation rule is a matter of policy rather than research, the validation or non-validation of the instrument review is based on the extent of physician agreement. The most stringent decision rule for validation of non-acute care requires 100% agreement by physicians that care is non-acute. Conversely, the most lenient confirmation rule requires at least one physician (one out of three) to corroborate non-acute care. For purposes of this project it is argued that there is some support for the criteria when at least one physician agrees that care could have been performed in other than an acute-care setting. On the other hand,

in a "real life" review (e.g., pre-certification and/or concurrent utilization review or retrospective appeals process), the majority rule would typically prevail. The validity of the criteria can, therefore, be assessed on minimum (lenient), medium (two out of three) or maximum (stringent) standards.

Three measures of reliability were computed for the sub-sample of 99 charts selected for the reliability trials: An overall measure of agreement, and two specific measures of agreement, specific acute and specific non-acute reliability. The kappa statistic was used to judge the overall reliability of the instruments. A significant kappa was judged as sufficient to suggest that an instrument was performing beyond chance agreement rates. The crucial aspect of the instruments under investigation is the ability of the criteria to be consistent on findings of non-acute care.

Once the reliability of an instrument was established, the validity of the criteria was determined in a manner similar to that used for the reliability trials. Criteria found to be reliable (specific non-acute reliability in excess of 50%) would be validated on a sub-set of 175 cases. Overall and specific non-acute agreement between the criteria and expert physician judgment is used to assess the validity of the instrument(s).

Each chart was reviewed by three (3) physicians from each organizational (FFS and HMO) setting. Validity was established by applying a validation rule based on the number of physicians agreeing with instrument findings (e.g., one or more, two or more or three out of three physician agreement). Estimates of non-acute care (see estimation results, Chapter III, Section III) are validity adjusted to represent physician estimates of non-acute care as well as

the degree to which the instrument(s) over or underestimate non-acute care. The construction of the hospital sample and the controlled selection process is now presented.

Sample

In this study, general, short-term, non-psychiatric, non-pediatric, non-federal, acute care facilities in southeast Michigan constitute the study's hospital frame. In this study, we use a probability proportionate to size sampling technique and the controlled selection process (Hess, 1975) to generate a sample of hospitals from the hospital frame representing the acute care facilities in Michigan. The sample was stratified by geographic area, bed size and occupancy rate with the probability of hospital selection proportionate to the number of admissions to the facility. [Teaching status was not used as a stratification variable since 68% of the teaching hospitals were already represented in the sample with certainty (self-representing).] Each of the cells in the controlled selection process represents an interaction of the three stratification variables. Since the numbers within a cell represent the admissions to that cell, the probability of choosing a cell increases as the admissions to the cell approach the critical number distinguishing self-representing (13,000 + admissions per year) from non-self-representing hospitals (less than 13,000 admissions per year). Controlled selection identifies the number of hospitals to be drawn from a particular cell. This selection was done by a random drawing from among the hospitals within a cell. The larger hospitals (those with more annual admissions) within a cell are more likely to be chosen for inclusion in the sample frame.

Table II - 1 presents comparisons between the hospital frame and the sample that was generated by the controlled selection process. The original sample of hospitals, from the universe of hospitals, was 80 with 38 hospitals self-representing (included in the sample with certainty), i.e., the number of admissions to these hospitals exceeded 13,000 admissions during the year 1983. Since the original sample of cases selected by the controlled selection process was self-weighting, there was no need to weight individual hospitals for the purpose of estimating non-acute care. As will be discussed in the following sections, resources constraints and voluntary participation in the study limited our final hospital sample size to $\underline{N}$ = 21 from southeast Michigan (population of hospitals = 80).

Insert Table II - 1 About Here

Record (Case) Selection

The selection of the admission sample from within the frame of admissions to facilities in Michigan was completed in the following manner:

(1) Prior to sampling the records, the data base was purged of the following categories of admissions:

* Obstetrics: ICD-9-CM Codes 630.0 - 676.9;
* Psychiatric: ICD-9-CM Codes 290.0 - 319.0;
* Children under the age of 18.

(2) Incomplete records, or records containing errors were excluded from the study resulting in 731,859 cases remaining in the sample. This number was distributed across the three payor categories (all other sources of payment, i.e., self-pay and commercial insurance were not included in the study) as follows:

* Blue Cross Admissions: 376,816 (51.49%);
* Medicare Admissions: 254,871 (34.83%);
* Medicaid Admissions: 100,172 (13.69%).

From this universe of admissions, 2,500 records were to be selected and reviewed. The sample was self weighting within and between self-representing and non-self-representing hospitals.

Response Rate and Over-Sampling

When field work began, it became clear that all the hospitals selected would not participate in the study and that financial resources would allow approximately 1,200 reviews to be performed. A decision was made to over sample the number of records in hospitals that agreed to participate in an attempt to attain a revised goal of 1,250 case reviews. Of the 40 hospitals in southeast Michigan selected via the controlled selection process to participate in the study, 21 hospitals eventually agreed to take part in the research. The 21 hospitals in the study provide estimates of non-acute care limited to the southeast Michigan area (see Map II - 1).

Table II - 2 compares the sample of 21 hospitals in the study with the hospitals in By-Law District II originally identified by the controlled selection process and for which records were available for review. One hospital had to be eliminated from the study because the records were on microfiche and were difficult for the review staff to read. The final hospital sample numbered 21, with 1,266 reviews performed over the 21 hospitals. As shown in Table II - 2, the proportion of responding hospitals in a variety of the subgroups are quite close to the proportions in the full sample. Thus, despite the low hospital response rate (50%), the respondent hospitals appear to be reasonably representative of all District II hospitals. The methodology involved in the selection of the day sample is presented next.

Insert Table II - 2 About Here

Day Sample

Within an admission, a maximum of seven days was sampled for review. In admissions with a length of stay of seven (7) days or less, all days were reviewed. For admissions with lengths of stay exceeding seven days, a maximum of five (5) days was sampled between the admission and discharge dates. The admission and the day before the actual discharge day were included with certainty in the day sample.

Data Collection

By May 31, 1985, project RN review staff, using the AEP and the SMI, completed on-site reviews of 1,266 charts over the 21 hospitals in the study. (However, both the AEP and the SMI were applied to 1,177 identical cases.) A description of the charts reviewed by payor source is shown in Figure II - 1. On average, the number of charts reviewed per hospital was 60. A minimum of 25 to a maximum of 112 charts were reviewed in each hospital, depending on the annual 1983 admissions to a given facility. Reliability and validity methods and their applicability to estimation are presented in the following sections.

Insert Figure II - 1 About Here

Inter-Rater Reliability

Inter-rater reliability is the extent to which the findings of one independent nurse reviewer are replicated by another independent nurse reviewer. Ninety-nine (99) charts, 7.82% of the 1,266 cases in the sample were subjected to reliability trials. Three types of reliability were calculated: (1) an overall measure of agreement; (2) specific acute reliability; and (3) specific non-acute reliability. The overall measure of reliability reflects agreement on combined acute and non-acute findings. The kappa statistic is used to determine, by statistical tests of significance, whether agreement is beyond that which might be reached by chance alone. Thus, a significant kappa indicates better than random agreement (significantly different from chance agreement) when measuring overall agreement.

The main focus in this study is specific non-acute reliability. The comparative strength of an instrument is judged not only by the degree of overall reliability, but, more importantly, by the degree of specific non-acute reliability. Instruments biased in favor of acute care need very little agreement on non-acute care to receive extremely high overall rates of agreement and relatively high kappa statistics. Thus, the relationship of specific non-acute reliability and the significance of the kappa statistic will be used to determine reliability.

Inter-rater reliability is one method to determine agreement. The problem with inter-rater reliability is that one is not sure whether the reliability of the instrument, or the reliability of the rater, or some combination of both is being measured. To control for these uncertainties, intra-rater reliability measures were also computed. Intra-rater reliability measures the consistency with which the criteria are applied, while holding constant rater differences. Three months after the initial reviews, raters were asked to re-evaluate the charts they had previously reviewed.

In conjunction, inter- and intra-rater reliability provide a measure of the performance of the instrument in consistently identifying non-acute care. A high score on both is indicative of an extremely reliable instrument. A high intra-rater with a low inter-rater score is indicative of a weak instrument. The reliability measures were computed overall and individually on medical and surgical admissions, medical days, and days within surgical admissions.

Reliability measures were computed as follows:

$$\text{Overall Reliability} = \frac{\text{\# of Cases Found Acute and Non-acute By Both Raters}}{\text{Total Number Of Cases}}$$

$$\text{Specific Non-acute Reliability} = \frac{\text{\# of Cases Both Raters Found Non-acute}}{\text{\# of Cases Either Rater Found Non-Acute}}$$

$$\text{Specific Acute Reliability} = \frac{\text{\# of Cases Both Raters Found Acute}}{\text{\# of Cases Either Rater Found Acute.}}$$

$$\text{Kappa Statistic } * = \frac{\text{\% Observed Agreement - \% Expected Agreement}}{\text{1 - \% Expected Agreement}}$$

Establishing the reliability of an instrument is a necessary but not sufficient pre-condition for its use. A reliable instrument must also be validated, or proven to be accurate. In our next section, we describe our approach to validating the instruments.

---

* The kappa statistic (Cohen, 1960) measures agreement and not association. Unlike measures of association, kappa explicitly adjusts for the amount of agreement occurring due to chance alone. The kappa statistic (K) may be tested for being significantly different from zero (0) by means of a simple t-test using the normal distribution which may be expressed as a probability with varying degrees of confidence. A kappa of zero (0) indicates that results are due entirely to chance. A kappa of one (1.00) indicates that agreement is due entirely to the criteria-set's ability to consistently identify acute and non-acute care. A kappa statistic between 0 and 1 indicates varying degrees of chance/non-chance association. A kappa statistic significantly different from zero (0) indicates that agreement is significantly beyond that expected by chance association.

Validity

Validity was established by comparing instrument findings with the clinical judgments of a panel of physician reviewers. Each of the 175 cases (see List II - 1) in the validity sample (see validation results section for a complete description of the validity sample) was reviewed by three physicians from a FFS setting and three different physicians from an HMO setting. Internists reviewed only medical cases. Surgeons primarily reviewed surgical cases, although some medical cases with significant surgical procedures were reviewed by surgeons.

Insert List II - 1 About Here

Selection of the physician panel was made by the principal investigator (Ira Strumwasser). One panel was composed of 11 physicians from Henry Ford Hospital, a predominantly fee-for-service, high quality, teaching facility with salaried staff physicians. The second panel of physicians was from Kaiser Permanente Medical Group of Oakland, California, a high quality, well known group model HMO.

The validity of the instrument was tested against three rules. The liberal rule required at least one of the three physicians to agree with the instrument finding that care was non-acute. The moderate rule required at least two of the physicians to agree with the instrument finding; while the conservative rule required all three physicians to agree that care was non-acute. It was expected that specific non-acute validity would fall as the instrument findings are compared to successively more stringent agreement (validation) rules.

As the rule is strengthened, from liberal to conservative, it should become harder to establish physician agreement with instrument findings of non-acute care. (This would require complete and consistent agreement among physicians.) Which specific rule is the right rule, and what fraction of the care is validated is a policy, and not a research, decision. It depends on what level of non-acute care is determined to be acceptable, since there will never be **complete** agreement on what cases are appropriate cases for hospitalization. The kappa statistic is once again used to measure the degree to which instrument/physician agreement is obtained beyond chance. As in the case of reliability, the focus is on specific non-acute validity or the agreement between the criteria and physicians in identifying non-acute care.

Estimates

Estimates presented in the pages that follow are validity adjusted (see Estimates of Non-acute Care, Chapter III, section III). Rates of non-acute care derived from application of the criteria are multiplied by specific non-acute validity coefficients. The resultant rate indicates the fraction of care in these hospitals that would be judged to be non-acute based on physician assessment. Estimates of non-acute care within hospitals is the ratio of the number of admissions found by the instrument to be non-acute in relation to the total number of admissions to that facility.

Estimates are provided for medical/surgical admissions as well as for medical and surgical days. Estimates of the fraction of non-acute days within acute and non-acute admissions are also provided. Estimates were computed by sex, age and payor. Validity adjusted estimates of non-acute care represent a

range of non-acute care values. The highest rates are based on the liberal agreement rule and the lowest based on the more stringent conservative validation rule. Finally, validity coefficients and rates of non-acute care for both FFS and HMO physicians are presented.

TABLE II - 1

Comparison Of The "Controlled Selection" (CS) Sample (N=80) With The Universe Of Acute-Care, General Hospitals In Michigan's Lower Peninsula (N=183)

HOSPITAL CATEGORIZATION	CS SAMPLE (%) (1)	UNIVERSE (%) (2)
TOTAL	80 (100)	183 (100)
SELF REPRESENTING	38 (47.50)	38 (20.76)
NON-SELF REPRESENTING	42 (52.50)	145 (79.23)
SMSA	66 (82.50)	122 (66.67)
NON-SMSA	14 (17.50)	61 (33.33)
TEACHING	26 (32.50)	29 (15.84)
NON-TEACHING	54 (67.50)	154 (84.15)
TAX STATUS=1 (Non-Profit)	51 (63.75)	128 (69.94)
TAX STATUS=2, 3 (Church, County)	29 (36.25)	55 (30.05)
BED SIZE=1 (0-149)	15 (18.75)	96 (52.45)
BED SIZE=2 (150-249)	16 (20.00)	33 (18.03)
BED SIZE=3 (250 >)	49 (61.25)	54 (29.50)
OCCUPANCY RATE=1 (0-59.9%)	12 (15.00)	65 (35.51)
OCCUPANCY RATE=2 (60-79.9%)	36 (45.00)	76 (41.53)
OCCUPANCY RATE=3 (80% >)	32 (40.00)	42 (22.95)
BED SIZE=1 (0-149)		
OCCUPANCY RATE=1 (0-59.9%)	9 (11.25)	59 (32.24)
OCCUPANCY RATE=2 (60-79.9%)	6 (7.50)	33 (18.03)
OCCUPANCY RATE=3 (80% >)	0 (0.00)	4 (2.18)
BED SIZE=2 (150-249)		
OCCUPANCY RATE=1 (0-59.9%)	2 (2.50)	5 (2.73)
OCCUPANCY RATE=2 (60-79.9%)	11 (13.75)	21 (11.48)
OCCUPANCY RATE=3 (80% >)	3 (3.75)	7 (3.83)
BED SIZE=3 (250 >)		
OCCUPANCY RATE=1 (0-59.9%)	1 (1.25)	1 (0.55)
OCCUPANCY RATE=2 (60-79.5%)	19 (23.75)	22 (12.01)
OCCUPANCY RATE=3 (80% >)	29 (36.25)	31 (16.94)

(1) Hospitals initially sampled (N=80) and selected for participation in the study based on the Controlled Selection Process.
(2) The total population of short-stay, acute-care, general hospitals located in Michigan's lower peninsula.

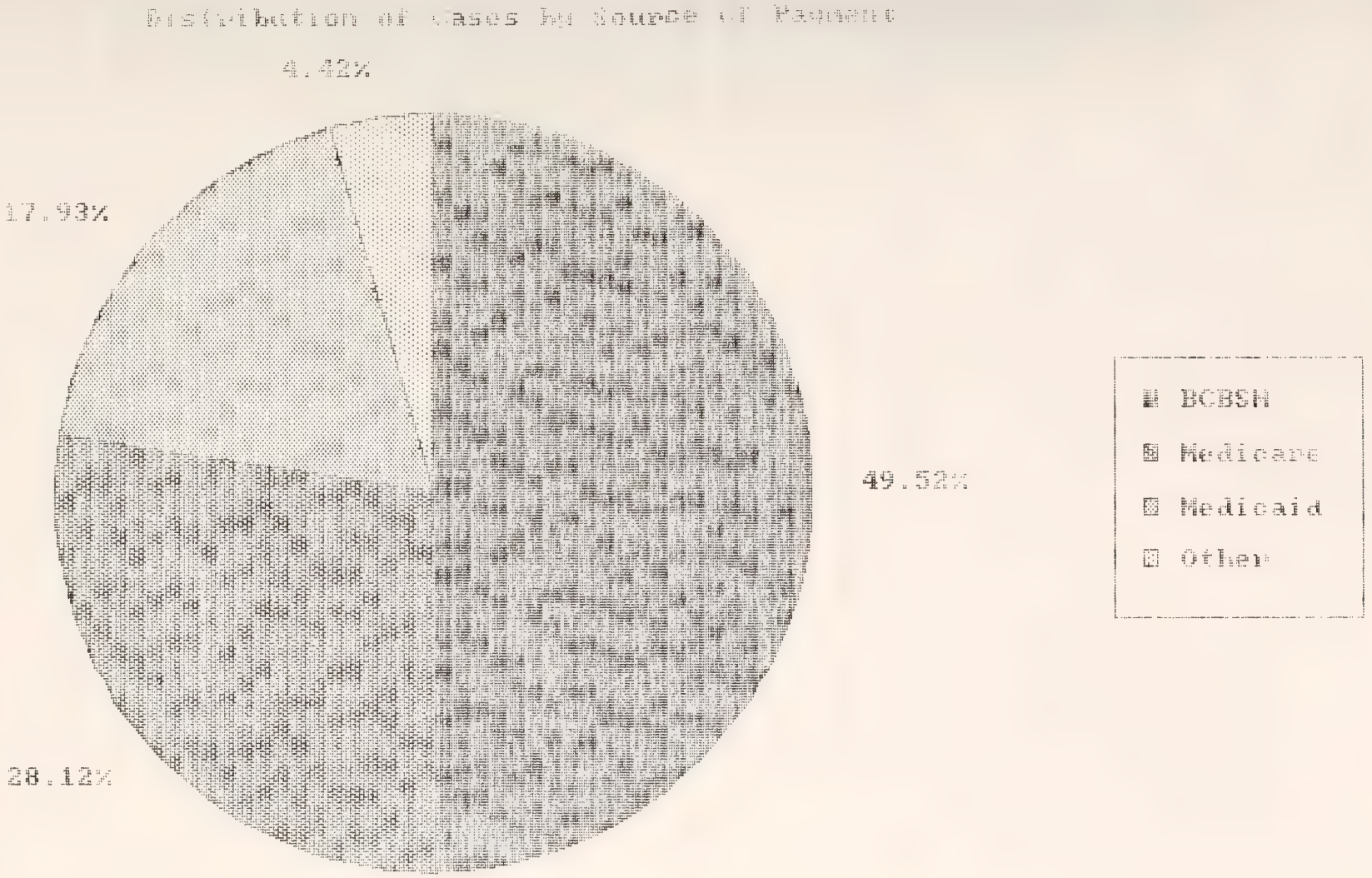
TABLE II - 2

Comparison Of CS Sampled Hospitals From All By-Law Districts That Agreed To Participate, Of Hospitals in By-Law DISTRICT II

HOSPITAL CATEGORIZATION	(1) SAMPLE (%)	(2) DISTRICT II (%)
TOTAL	21 (100.00)	40 (100.00)
SELF REPRESENTING	13 (61.90)	22 (55.00)
NON-SELF REPRES.	8 (39.10)	18 (45.00)
SMSA	21 (100.00)	40 (100.00)
NON-SMSA	0 (0.00)	0 (0.00)
TEACHING	8 (38.09)	12 (30.00)
NON-TEACHING	13 (61.90)	28 (70.00)
TAX STATUS=1 *	14 (66.67)	27 (67.50)
TAX STATUS=2, 3 *	7 (33.33)	13 (32.50)
BED SIZE=1	1 (4.76)	4 (10.00)
BED SIZE=2	5 (23.80)	7 (17.50)
BED SIZE=3	15 (71.42)	29 (72.50)
OCCUPANCY RATE=1	1 (4.76)	2 (5.00)
OCCUPANCY RATE=2	5 (23.80)	14 (35.00)
OCCUPANCY RATE=3	15 (71.42)	24 (60.00)
BED SIZE=1 (0-149)		
OCCUPANCY RATE=1	1 (4.76)	2 (5.00)
OCCUPANCY RATE=2	0 (0.00)	2 (5.00)
OCCUPANCY RATE=3	0 (0.00)	0 (0.00)
BED SIZE=2 (150-249)		
OCCUPANCY RATE=1	0 (0.00)	0 (0.00)
OCCUPANCY RATE=2	3 (14.28)	4 (10.00)
OCCUPANCY RATE=3	2 (9.52)	3 (7.50)
BED SIZE=3 (250 >)		
OCCUPANCY RATE=1	0 (0.00)	0 (0.00)
OCCUPANCY RATE=2	2 (9.52)	8 (20.00)
OCCUPANCY RATE=3	13 (61.90)	21 (52.50)

(1) The 21 hospitals from southeast Michigan included in the study.
(2) Population of hospitals located in By-Law District II, selected via controlled selection.
(*) Tax Status 1 = non-profit; Tax Status 2 & 3 = Church, City, County, State

Figure II - 1


Distribution of Cases by Source of Payment
4.42%
17.93%
49.52%
28.12%
BCBSM
Medicare
Medicaid
Other

LIST II - 1

Charts Duplicated For Intra- And Inter-Rater Reliability And Physician Reviews By Major Diagnostic Categories (ICD-9-CM Nomenclature)

DIAGNOSIS	ICD-9-CM	No.	Percentage
Infectious and Parasitic Diseases	(001-139)	4	2.28
Neoplasms	(140-239)	19	10.85
Endocrine, Nutritional and Metabolic Diseases and Immunity Disorders	(240-279)	8	4.57
Diseases of the blood and blood forming organs	(280-289)	0	0.00
Mental Disorders	(290-319)	1	0.57
Diseases of the Nervous System and Sense Organs	(320-389)	7	4.00
Diseases of the Circulatory System	(390-459)	34	19.42
Diseases of the Respiratory System	(460-519)	10	5.72
Diseases of the Digestive System	(520-579)	25	14.28
Diseases of the Genitourinary System	(580-629)	14	8.00
Complications of Pregnancy, Childbirth and the Puerperium	(630-676)	0	0.00
Diseases of the Skin and Subcutaneous Tissue	(680-709)	5	2.85
Diseases of the Musculoskeletal System and Connective Tissue	(710-739)	27	15.42
Congenital Anomalies	(740-759)	0	0.00
Certain Conditions Originating in the Perinatal Period	(760-779)	0	0.00
Symptoms, Signs and Ill-Defined Conditions	(780-799)	9	5.14
Injury and Poisoning	(800-999)	11	6.28
Missing	(0)	1	0.57
TOTAL		175	100.00

CHAPTER III

RESULTS

Reliability

RELIABILITY

Most methodological studies on the appropriateness of hospitalization have had difficulties in obtaining reliable measurements of non-acute care. Reviewers tend to obtain moderately high levels of overall agreement on the appropriateness of care. However, with the exception of the results reported by SysteMetrics (1984) and Gertman and Restuccia (1981), previous methods for determining the appropriateness of hospitalization were found to be unreliable. The methodological importance of inter-rater reliability in the assessment of inappropriate hospitalization is critical. Without high inter-rater reliability coefficients, the validity of an instrument cannot be determined. Therefore, a brief review and discussion of the issues regarding rater reliability in the use of the AEP and SMI is presented.

In the early studies, a large proportion of the overall agreement was due to chance since most care was judged to be appropriate by reviewers. For example, if two reviewers believe that 90% of hospital care was appropriate, by chance alone, the overall agreement rate would be 82% [i.e., (.9 x .9) + (.1 x .1)]. To illustrate this point, consider the associations that may be expected by chance in the research conducted by SysteMetrics (1984) and Gertman and Restuccia (1981). Restuccia et al. (1984) report the AEP finds approximately 19.1% of all admissions and 20% of all days of care within otherwise acute admissions to be non-acute. Similarly, SysteMetrics reports 5.7% of the admissions and 7.8% of the days in admissions deemed to be acute were found to be non-acute. Were reviewers to, a priori, expect to find these rates of non-acute care, the overall agreement rates for the respective instruments in these studies, due to chance association alone, would be as follows:

AEP Admissions = (.809 x .809) + (.1 x .1) = 66.45%;

AEP Days = (.800 x .800) + (.1 x .1) = 65.00%;

SMI Admissions = (.943 x .943) + (.1 x .1) = 89.92%;

SMI Days = (.920 x .920) + (.1 x .1) = 85.64%.

Thus, in the research conducted by Restuccia et al. and SysteMetrics, overall reliability coefficients for admissions which exceed 66% for the AEP and 90% for the SMI would be beyond those expected by chance association. Similarly, overall reliability coefficients for days which exceed 65% for the AEP and 86% for the SMI would be due to other than chance association.

Specific Non-Acute Agreement

The overall agreement, or the rate of agreement on all cases reviewed, whether judged acute or non-acute is of general importance. However, specific agreement, or the rate of agreement on cases determined to be non-acute, is the critical reliability comparison. Without reasonable levels of specific non-acute agreement among reviewers, there cannot be valid or reliable estimates of non-acute hospital care. It is impossible to determine the degree to which estimates of non-acute care is attributable to random chance, reviewer bias, or actual inappropriateness without a highly reliable and valid measure of non-acute care.

Two of the principal statistical measures of the reliability of methodological approaches with seldom occurring events are: (1) specific agreement, and (2) the kappa statistic. Specific non-acute agreement is the ratio of the number of cases deemed non-acute by both raters to the number of cases judged non-acute by either of the raters (see Table III - 1). The kappa statistic (Cohen, 1960) measures agreement and not association. Unlike measures of association, kappa explicitly adjusts for the amount of agreement occurring due to chance alone. Kappa may be expressed as:

$$K = \frac{\text{\% observed agreement - \% expected agreement}}{\text{1 - \% expected agreement.}}$$

Insert Table III - 1 About Here

TABLE III - 1

Inter-Rater Agreement on Necessity
of Hospitalization

		Rater 1	
		Acute	Non-Acute
Rater 2	Acute	a	b
	Non-Acute	c	d

Overall agreement = $\frac{a + d}{a+b+c+d}$ x 100.

Specific non-acute agreement = $\frac{d}{b+c+d}$ x 100.

The kappa statistic (K) may be tested for being significantly different from zero (0) by means of a simple t-test using the normal distribution which may be expressed as a probability with varying degrees of confidence. A kappa of zero (0) indicates that results are due entirely to chance. A kappa of one (1.00) indicates that agreement is due entirely to the criteria-set's ability to consistently identify acute and non-acute care. A kappa statistic between 0 and 1 indicates varying degrees of chance/non-chance association. A kappa statistic significantly different from zero (0) indicates that agreement is significantly beyond that expected by chance association.

We now turn to a review of the previous results of reliability trials conducted by instrument developers and reported in the literature by SysteMetrics (1984), Gertman and Restuccia (1981) and Restuccia et al. (1984).

Outcome Reliability As Reported By Instrument Developers

A close review of the research reports indicate that SysteMetrics and Gertman and Restuccia report very similar reliability coefficients between the AEP and the SMI. According to research by Gertman and Restuccia (1981) and SysteMetrics (1984), the AEP and the SMI compare favorably for overall reliability on days of stay (see Table III - 2). Gertman and Restuccia (1981) report corrected overall inter-rater reliability of between 91.8% and 94.3% for days of stay (kappa statistic significant beyond $p < .0001$). SysteMetrics (1984) reports an (uncorrected) overall reliability on days of stay of 98.7% (kappa statistic = 79.8, $p < .001$). (The original version of the AEP did not evaluate the appropriateness of the admission. However, a subsequent version

of the AEP, used and evaluated in the present research, was designed to evaluate the appropriateness of both medical and surgical admissions.) Thus, while reliability ratios for admission reviews for the AEP are not available, the SMI reports impressive overall reliability for both admission and day reviews. The overall reliability for day reviews, as reported by Gertman and Restuccia (1981) and SysteMetrics (1984), for both the AEP and SMI are greater than 90% with statistically significant kappas.

Insert Table III - 2 About Here

However, when specific agreement is examined, the reliability for both instruments drops substantially. Specific agreement on admissions reported by SysteMetrics is 67.35% (specific agreement for the AEP admission reviews were not available). Specific agreement for days of stay reported by Gertman and Restuccia for the AEP are between 73.1% and 79.3%; while the specific (uncorrected) agreement for days of stay for the SMI, reported by SysteMetrics, is 62.5%. While comparative statistics for specific agreement between the AEP and SMI for admissions are not available, research conducted by the instrument developers indicate that the AEP is slightly more reliable than the SMI in assessing non-acute days of care.

While SysteMetrics (Final Report, pg. IV 32 - IV 33, 1984) reports overall inter-rater reliability coefficients on admissions to be 97.6% (kappa = .76, $p < .001$) and the reliability coefficient for days as 98.7% (kappa = .80, $p < .001$), these high rates of overall agreement are due primarily to the relatively low rates of non-acute admissions (between 4.83% and 5.64%) and days of care (between 2.97% and 3.61%) (see Tables III-3 and III-4). As

indicated in Table III - 2, specific agreement for the SMI on admissions is 67.35% and the specific agreement for days of care is 62.5%.

Insert Table III - 3 and III - 4 About Here

Gertman and Restuccia (1981) report that overall agreement between pairs of raters on days range from 91.8% to a high of 94.3% ($p < .0001$, kappa statistic). While Gertman and Restuccia report that specific reliability for days of stay are between 73.1% and 79.3% (for pair-wise comparisons), it is important to note that these reliability coefficients are based on corrected data i.e., reviews corrected by Gertman and Restuccia. In a footnote on page 865 [Gertman and Restuccia, (1981)], the authors note that uncorrected reviews (i.e., those not checked and altered by Gertman and Restuccia) were between 84% and 91% ($p < .01$) for overall reliability and between 58% and 72% for specific day agreement. Similarly, Rishpon, Lubasch and Epstein (1986), in a recent and brief research paper, report the specific non-acute day reliability of the AEP to be between 52.4% and 57.9%, ($p < .008$). Thus, based on uncorrected reliability data (e.g., the reliability we might expect in actual use of the criteria-set), the AEP and SMI appear to report reliability coefficients, whether overall or based on specific reliability, that are quite similar. While the corrected reliability indicates the theoretical absolute maximum reliability of the AEP, for purposes of comparison, the appropriate reliability data comparing the AEP with the SMI should either be corrected or uncorrected data. Since SysteMetrics reports only uncorrected reliability data, and since we are interested in reliability on actual rather than theoretical application, we shall use these uncorrected data. Comparable data for

admission reviews are not available since, at that point in time, the AEP did not assess the appropriateness of admissions and the authors of this report (Strumwasser and Paranjpe) have been unable to obtain current admission reliability data on the AEP.

We now turn to the comparative empirical analysis of the reliability of the AEP and the SMI. In addition to providing an analysis of the comparative reliability of admission and day reviews for adult medical care, an analysis of the reliability of the newly developed instrument called the Surgical Appropriateness Evaluation Protocol (SAEP) is also provided.

Reliability Methods

To assess the reliability of the AEP and the SMI, 99 charts were selected from 13 hospitals. An average of eight (8) records were selected for reliability review from each hospital, a 7.82% sample of all cases reviewed over the 21 hospitals included in the study. Since these records would be used for both the reliability trials and subsequent validity assessment, they were purposely chosen to reflect instances in which the AEP and the SMI: (1) agreed that an admission was acute, (2) agreed that an admission was non-acute, and, (3) instances in which the two instruments disagreed on the appropriateness of the hospitalization. Table III - 5 indicates the distribution of cases on this admission agreement/disagreement model ($n = 99$) selected for the reliability reviews while Tables III - 7 and III - 8 present the total number of randomly

---

Insert Tables III - 5, III - 7 and III - 8 About Here

---

selected cases reviewed for the estimates of non-acute care (N = 1,173). Since one of the primary reasons for conducting the study was to assess the validity of each criteria set, especially on non-acute findings, the largest number of cases selected for reliability trials and subsequent validity assessments were from the cells in which the two instruments disagreed. The most frequent direction for disagreement between the two criteria-sets was the instance in which the AEP found the admission to be non-acute while the SMI found the admission to be acute. The majority (68.69%) of sampled cases chosen for reliability review were selected from this intersection (see Table III - 5). Because of this sampling procedure, we will provide both unweighted and subsequently weighted reliability coefficients. (See Technical Note on weighted reliability, Attachment D.) Since it was also important to determine whether or not instruments were reliable and valid in instances in which they both agreed that care was either acute or non-acute, a small sample of charts was selected from these cells. In general, the selection of cases for duplication and subsequent re-review was based on the following rules: (1) Of all records reviewed within a given hospital, one case was randomly selected in which both instruments agreed the admission was acute; (2) Of all records reviewed within a given hospital, two cases were randomly selected in which both instruments agreed the admission was non-acute; (3) One case, should it be found to exist, was randomly selected in which the SMI found the admission to be non-acute while the AEP found the admission acute (this infrequently occurred since the SMI is more likely to find care to be acute than the AEP); (4) At least five cases in which the two instruments disagreed on the appropriateness of the admission and where the AEP found the admission non-acute while the SMI found the admission acute were randomly selected for inclusion

in the reliability sample. This sample constitutes the subset of 99 cases (Table III - 5) selected for the reliability trials.

It should be noted that the reliability subset sample is a purposefully biased sample. The reliability subset is biased toward cases in which the admission is found to be non-acute by one (the AEP) or both instruments. For example, 79.79% of the admissions were non-acute according to the AEP while 14.14% of the admissions were non-acute according to the SMI. Thus, the unweighted reliability coefficients are likely to reflect extremely low agreement for instruments that do poorly on non-acute assessments. Since the primary purpose of these criteria-sets is to provide reliable and valid assessments of non-acute care, this biased subset seems appropriate. It is also possible that cases in which the admission is considered to be non-acute by the AEP and acute by the SMI are among the more difficult cases to evaluate thereby testing the performance of the protocols under the most rigorous conditions. However, for population estimates, reliability will vary as a function of the extent of acute and non-acute care reported by each criteria-set. Thus, to approximate expected population reliability coefficients, weighted reliability coefficients will also be provided as a technical note (see Attachment D).

The initial review of charts was performed on-site in each of the 21 hospitals participating in the study. In total, 1,266 unique in-hospital chart reviews using each instrument were performed. (However, both instruments were applied to only 1,173 identical cases.) These reviews took place between February 5, 1985 and May 31, 1985. The 99 charts selected for inter-rater and intra-rater reliability reviews were from the first 13 hospitals in which charts were reviewed. All reviews for the 99 charts chosen for the reliability assessment

took place during February, March and April, 1985.

Two forms of rater reliability were examined. Inter-rater reliability reflects the consistency of reviews between two (2) different raters. Intra-rater reliability reflects the consistency of reviews within each rater. Table III-6 illustrates a hypothetical relationship and interpretation of the interaction between the inter- and intra-rater reliability.

Insert Table III - 6 About Here

Inter-rater reliability assesses the reliability of reviews between raters who may have different levels of understanding, training, and ability, and between reviewers who may interpret the application of the instruments differently. Intra-rater reliability holds individual rater differences constant. It asse-sses the degree of replicability while holding idiosyncratic rater differences constant. As suggested in Table III - 6, for the application of an instrument to be considered highly reliable, both inter- and intra-rater reliability should be high. Unreliable application of an instrument is reflected by both low inter- and low intra-rater reliability. Contrasting inter- and intra-rater reliability provides information on the source or cause of low reliability, e.g., internal consistency/inconsistency (intra-rater reliability) versus external consistency/inconsistency (inter-rater reliability).

RELIABILITY RESULTS

SUMMARY

The results of the reliability trials between the Appropriateness Evaluation Protocol and the Standardized Medreview Instrument, compiled by this research, indicates that the AEP is reliable, while the SMI is an unreliable instrument for identifying non-essential hospital care.

In general, this research confirms the inter-rater reliability coefficient for specific non-acute hospitalization reported by Gertman and Restuccia (1981). The specific non-acute reliability reported by SysteMetrics (1984) of 62% - 67% was not substantiated. Overall, the AEP attained specific non-acute (unweighted) reliability coefficients of 75% ($p < .01$) for admissions. The SMI's specific non-acute reliability was 10%. Similarly, for the DOS model, application of the AEP resulted in a specific non-acute (unweighted) reliability coefficient of 81% ($p < .01$). SMI produced a specific non-acute (unweighted) reliability coefficient of 35% ($p < .01$).

Since the SMI reliability coefficients were low, additional admission reviews were performed using Professional Review Organization (PRO) auditors (see SMI Reliability Addendum, Attachment E). The (unweighted) admission reliability of the SMI using a reviewer employed by SysteMetrics and two RN reviewers employed by an Indiana PRO (which uses the SMI for hospital audits) is higher (moderately good) than the coefficients of agreement obtained by BCBSM reviewers applying the SMI. However, the coefficients of agreement obtained by SysteMetrics and PRO reviewers are substantially lower than those obtained using BCBSM reviewers applying the AEP.

(UNWEIGHTED) INTER-RATER RELIABILITY

Admission Review

The AEP and the SMI have roughly equivalent overall reliability coefficients for admission reviews. However, the data indicates that the AEP reliably identifies admissions found to be non-acute while the SMI does not reliably identify admissions found to be non-acute. [The supporting Tables for the remainder of the Reliability Section (Tables III - 9 through III - 89) may be found in Appendix A, Book II.]

As indicated in Table III - 9, the overall inter-rater reliability for admission review is 79% for the AEP and 73% for the SMI. However, specific non-acute inter-rater reliability for the AEP is 75% and 10% for the SMI. The kappa statistic for the AEP is significantly different than zero (chance) (kappa = .47, $p < .01$).

Insert Table III - 9 About Here

These overall inter-rater reliability coefficients on admission review are gross measures. They combine assessments of medical and surgical admissions and average reliability over four different raters. Distinguishing between medical and surgical admissions is important for both substantive and methodological assessment reasons. Substantively, it may be more or less difficult to assess the appropriateness of surgical than medical admissions (or days of stay within medical versus surgical admissions). The indications and preparations for elective surgery admissions are typically based on intensity of service, severity of illness and timing of service. Previous work

(SysteMetrics, 1984; Restuccia et al., 1984) indicates that inappropriate admissions are more frequent for medical rather than surgical admissions. Lastly, the AEP specifically assesses the appropriateness of admissions with the Surgical Appropriateness Evaluation Protocol (SAEP) (see Attachment C) while the SMI does not claim to specifically assess the appropriateness of elective surgical procedures. (All elective surgical procedures are considered by the SMI to be acute.) Finally, it is of methodological and statistical significance, especially in interpreting the data, to know whether the relationship of multiple pairwise inter-rater reliability coefficients are equally consistent across all raters or reflect average coefficients of agreement produced by wide variation in rater reliability.

We now turn to separate analyses of the medical and elective surgical admission reviews. Of foremost interest is whether, within and between criteria-set, differences in reliability are found for these categories (medical versus surgical) of admissions.

Medical Admissions

As demonstrated in Table III - 12, inter-rater reliability coefficients for medical admissions mirror those reported earlier for all admissions. The overall rates of agreement for medical admissions are 79% for the AEP (kappa = .49, $p < .01$) and 71% for the SMI. However, the specific non-acute reliability coefficient for medical admissions using the AEP is 75% (K = .49, $p < .01$) and 10% for the SMI. The AEP reliably identifies non-acute medical admissions, while the SMI does not reliably identify non-acute medical admissions. We now turn to an analysis of the inter-rater reliability for elective surgery admissions.

Insert Table III - 12 About Here

Elective Surgery Admissions

The inter-rater reliability coefficient for the SAEP is 75% overall and 60% for specific agreement (see Table III - 15). Thus, the first independent assessment of the SAEP's ability to reliably identify non-acute care suggests that it is consistent approximately 6 out of 10 times (kappa = .53, $p < .01$). These findings should be considered tentative due to the relatively small number of elective surgery cases evaluated with the SAEP ($N = 24$) in the reliability sample. The overall reliability of 100% for the SMI is meaningless since reviewer instructions are that all Class I elective surgical admissions are acute (see Table III - 15).

Insert Table III - 15 About Here

These data have demonstrated that the AEP is a reliable instrument for identifying non-acute medical and surgical admissions while the SMI is unreliable in identifying non-acute admissions. Preliminary data suggests that the SAEP is a moderately reliable criteria-set for identifying non-acute surgical admissions.

We now turn to an analysis of the homogeneity of individual rater reliability coefficients. Confirmation of the aggregate rater findings will be confirmed if individual raters obtain consistent levels of rater reliability.

Multiple Pairwise Inter-Rater Reliability

The patterns of agreement (inter-rater reliability) between each team of reviewers sheds some light on the consistency or inconsistency of agreement between raters. The first set of inter-rater comparisons examines the pairwise reliability of each rater with the other three raters who, in combination, re-reviewed 100% of the cases originally reviewed by a rater in the hospital setting.

The data indicate that each AEP reviewer is, with one exception, uniformly consistent in their reliability on the appropriateness of admissions. As indicated in Table III - 18, individual overall reliability for AEP reviewers ranges from a low of 63% to a high of 84%. Specific non-acute reliability ranges from a low of 55% to a high of 82%. All AEP kappa statistics, with one exception, are significantly better than chance association ($p < .01$). These data indicate that the high overall and specific non-acute reliability of the AEP is consistent among reviewers.

Insert Table III - 18 About Here

The data (see Table III - 23) also indicate that reviewers using the SMI are unable to obtain satisfactory levels of specific agreement. Although raters obtain moderately high rates of overall reliability, this is due to the fact that reviewers using the SMI infrequently find care to be non-acute. When one reviewer determines an admission to be non-acute, none of the other reviewers is able to find the admission to be consistently non-acute. The highest rate of specific non-acute reliability is obtained by SMI reviewer "A" (25%), with

a kappa not statistically different from zero (0) or chance. These data indicate that high AEP and low SMI reliability appears to be the result of consistently high and consistently low reliability coefficients across all AEP and SMI reviewers.

Insert Table III - 23 About Here

These data lend support to the initial aggregate findings that raters using the SMI obtain moderately high levels of overall reliability, primarily due to the fact that most care assessed with the SMI criteria-set is judged to be acute.

The AEP reliably identifies non-acute adult medical admissions while the SMI does not. This finding holds true for group and individual rater level assessments. Although the data on the SAEP is tentative due to the relatively small N, the reliability coefficients for non-acute admissions with the SAEP suggests that the SAEP may prove to be reliable. Additional trials using more cases are necessary before definitive conclusions on the SAEP are drawn.

These findings, in light of the research performed by SysteMetrics' research are surprising and in contrast to those reported in the literature (see the reliability critique by Blumenfeld, 1985 for a detailed critique of the Syste-Metrics' Research). These unexpected findings may be due to the relatively small number of admission comparisons (N = 98). In addition, the subsample of cases tending to be biased toward cases judged by raters using the AEP to be non-acute may explain our reliability findings for the SMI. These cases may simply represent more complex case-mix assessments than used in the develop-

ment of the SMI. To address these potential confounding factors, we present additional methodological and analytic data.

The first critique of the admission reliability analysis may be advanced as follows: The total number of admission observations is too few to result in conclusive findings (although the number of admissions on which this reliability analysis is based is similar to those on which Gertman and Restuccia and SysteMetrics based their original research). The total number of admissions reviewed (medical and surgical), for the reliability sample, may represent a sample which is too small to base firm conclusions regarding SMI reliability.

The arguments in favor of supporting our results and in rejection of these arguments are as follows: First, with an almost identical number of observations, reliability of the AEP and SAEP was able to be confirmed. Second, we shall present data on the day of stay (DOS) assessment of the AEP and SMI criteria-sets. Since an average of 3.67 randomly sampled days within each sampled admission was also chosen for reliability trials, we have a significantly increased sample (N = 363 days of care) on which to base our analysis and conclusions regarding the reliability of the AEP and SMI criteria-sets as it relates to the day of stay evaluation. To address both the issue of the number of observations and other methodological concerns, we had charts re-reviewed by the same reviewer who performed the original review (intra-rater reliability). Finally, we present weighted population reliability estimates to consider the potential biasing effect of our reliability sampling methods (see Technical Note, Attachment D).

Day Model Criteria-Set: Sampling Methodology

A random sample of days from each sampled chart was selected (see Table III - 5). The admission itself and the day before the actual day of discharge were sampled with certainty. A maximum of five (5) days between the admission and the day before discharge was randomly sampled for inclusion in the day reliability sample. The first day of stay was not included in the analysis since admission reviews, in general, assess the appropriateness of the admission and the first day. (This is not true for the SMI which provides an independent assessment of the admission and first day of stay.) In lengths of stay (LOS) six (6) days or less, all days are sampled. In LOS greater than six (6) days, a maximum of 4 days between the admission day and day before discharge was randomly sampled. DOS for reliability testing are based on the DOS within the admission reliability sample.

The data indicate that both the AEP and the SMI reach reliability levels for day reviews which are significantly better than chance association (see Table III - 28). However, the AEP performs better than the SMI. As indicated in Table III - 28, the overall reliability is 84% ($K = .58$, $p < .01$) for the AEP and 67% ($K = .27$, $p < .01$) for the SMI. Moreover, reviewers using the AEP obtain specific non-acute reliability in excess of 80%. Reviewers using the SMI verified findings of non-acute DOS 35%. Reliability coefficients for the DOS, for both instruments, are higher than those for admission reviews.

The data presented in Tables III - 31 and III - 36 represent individual comparisons of each rater with subsequent reviews by any member of the team. The results indicate an overall reliability coefficient ranging from a low of 80% to a high of 86% for the AEP DOS review. All kappas are significant $p < .01$.

Specific non-acute reliability for raters using the AEP ranges from a low of 76% to a high of 83%. The AEP appears to be a reliable criteria-set for identifying non-acute days of care.

Insert Tables III - 31 and III - 36 About Here

While three (3) out of four (4) kappas for the SMI day reviews are significant ($p < .01$), the overall and specific reliability coefficients are lower than those obtained by the AEP reviews. With the exception of the reliability coefficients for rater "A", the specific non-acute reliability coefficients for reviewers using the SMI are between 31% and 37%. These coefficients are 50% lower than those obtained by the reviewers using the AEP. In general, days found to be non-acute using the SMI criteria-set are verified by subsequent independent review in three (3) out of ten (10) instances. The specific non-acute reliability coefficients obtained by rater "A" (73%) appears to be an exception.

As indicated in Table III - 31, the overall and specific non-acute reliability coefficients for the AEP are homogeneous. These inter-rater reliability coefficients confirm that the AEP is a reliable instrument. The individual rater level analysis for raters using the SMI DOS criteria-set (see Table III - 36) does not, in general, provide confirmation of reliability. Reviewers obtain a specific non-acute DOS reliability coefficient (with the exception of rater A) ranging from 31% to 37%.

We now turn to a brief examination of medical and surgical days of stay.

Medical DOS

The data indicate that the AEP is a more reliable instrument for assessing the appropriateness of a DOS than the SMI. As indicated in Table III - 67, the specific non-acute reliability for medical DOS is 81% (kappa = .55, $p < .01$) for the AEP and 35% (kappa = .26, $p < .01$) for the SMI. The AEP specific non-acute reliability coefficient for medical DOS is 2.3 times greater than the SMI specific reliability coefficient.

Surgical DOS

The reliability of AEP DOS within elective surgery admissions is lower than the specific non-acute reliability coefficient for DOS within medical admissions (see Table III - 70). The specific non-acute reliability coefficient for AEP DOS within elective surgical admissions is 47%. Reliability is higher for medical than surgical DOS. It appears to be more difficult to agree on appropriate surgical DOS than on medical DOS using the AEP criteria-set.

The next section examines intra-rater reliability. Inter-rater reliability assesses the degree of agreement between two different raters. Intra-rater reliability assesses the degree of consistency within each rater. Intra-rater reliability holds constant idiosyncratic rater differences which may be due to a number of extraneous factors (e.g., understanding, knowledge, ability, orientation, etc.). Intra-rater agreement also allows an analysis of the degree to which inter-rater reliability for SMI rater "A" (the rater who obtained moderately good rates of inter-rater agreement) is substantiated (if rater "A" obtains statistically significant and high specific non-acute intra-rater reliability coefficients, then some support for the reliability of the SMI will be found to exist).

(UNWEIGHTED) INTRA-RATER RELIABILITY

Admissions

All reviewers performed the initial reviews, on-site, during February, March and April of 1985. Subsequent reviews were performed at a central location using high quality duplicate copies of the original chart, during May 1985. As indicated in Table III - 41, reviewers using the AEP criteria-set were able to obtain overall intra-rater reliability coefficients of 86% (kappa = .62, $p < .01$) and a specific non-acute intra-rater reliability coefficient of 83% for admission reviews. These data demonstrate that reviewers using the AEP criteria-set are consistent in their findings between and within raters. While reviewers using the SMI criteria-set obtain moderately high overall intra-rater reliability of 79% (K = .32, $p < .01$), the specific non-acute reliability is 29% for admissions. These data tend to provide additional evidence that the SMI is not a reliable instrument for identifying non-acute care. Admission reviews are unreliable between and within reviewers. The reliability coefficient for the SMI is better for intra-rater than for inter-rater reliability.

Insert Table III - 41 About Here

Tables III - 44 and III - 49 provide evidence which supports the reliability of the AEP. The reliability of the SMI (using intra-rater comparisons) is not supported. Raters using the AEP obtain statistically significant overall reliability coefficients for three (Raters "A", "B" and "C") of the four raters (kappa statistic = .78, .55, and .65, $p < .01$, respectively). The AEP

intra-rater specific non-acute reliability coefficient for admissions ranges from a low of 67% to a high of 89%. One SMI rater (rater "B") was able to obtain a significant kappa (kappa statistic = .29, $p < .01$). One SMI rater obtained a specific non-acute reliability coefficient of zero (0), two SMI raters obtained intra-rater reliability coefficients of 25 and 28%, while one rater (rater "A") obtained an intra-rater reliability coefficient on admissions of 50% (not significantly beyond that expected by chance association). These data indicate that both inter- and intra-rater specific non-acute reliability for AEP admission reviews are statistically significant. The specific reliability coefficient for the SMI admission reviews are low within and between raters. We now turn our attention to an analysis of intra-rater reliability for the day of care assessment.*

Insert Tables III - 44 and III - 49 About Here

---

* RNs were instructed to work at their own pace. Quality and not quantity of reviews was the overriding concern. Thus, RNs performed an unequal number of reviews. For AEP raters, the number of reviews was roughly equally divided among the four reviewers. However, for the SMI condition, Rater "B" performed substantially more reviews than SMI reviewers "A", "C" or "D". Thus, results are biased or weighted toward Reviewer "B". This may have resulted in lower reliability coefficients. Had the number of reviews been weighted toward Rater "A", then non-acute reliability may have been higher. However, both SMI intra-rater and inter-rater admission and day reliability coefficients were low or not significantly different than chance association for SMI reviewers, suggesting that the unequal weight biased toward Rater "B" was probably not a critical factor influencing low SMI rater reliability. In addition, when the number of charts for review were equally divided among the SysteMetrics and PRO reviewers, while SMI reliability was higher than in the field trials, it remained lower than the AEP field trial reliability (see Attachment E).

Days of Stay

As indicated in Table III - 54, overall intra-rater reliability coefficients for days is significantly greater than chance for both the AEP (kappa = .69, $p < .01$) and SMI (kappa = .45, $p < .01$) reviews. The specific non-acute intra-rater reliability coefficient for the AEP (85%) is higher than the specific non-acute reliability coefficient obtained with the SMI criteria set (47%). While SMI intra-rater reliability coefficients for days are higher than the inter-rater reliability coefficients (see Table III - 62), they do not reach levels obtained by the AEP review criteria (see Table III - 57).

Insert Tables III - 54 and III - 57 and III - 62 About Here

Specific non-acute intra-rater reliability for the AEP team ranges from a low of 74% to a high of 98% (rater "C") while the SMI team ranges from a low of 42% to a high of 64% (rater "A"). While all the raters on **both** the AEP and SMI teams obtained significant kappas ($p < .01$), the AEP team's kappas are about 33% higher than those of the SMI team.

These results confirm that AEP raters are consistent both between and within raters for admissions as well as days of care. The SMI is not reliable between reviewers for either days or admissions or within raters for admissions. There is some evidence in support of the SMI intra-rater reliability for days but these levels of reliability do not approach those obtained with the AEP. There is also some support which suggests that SMI Rater "A" tends to perform more reliable reviews than Raters "B", "C" or "D". Before concluding, we shall briefly turn to an examination of intra-rater reliability

for medical admissions and medical DOS.

Medical Admissions and DOS

The data indicate that intra-rater reliability coefficients confirm the inter-rater reliability findings. As indicated in Table III - 72, the intra-rater specific non-acute reliability coefficient for medical admissions is 84% for the AEP and 29% for the SMI. The AEP specific non-acute intra-rater reliability coefficient (84%) is roughly 2.9 times as great as the SMI specific non-acute reliability coefficient (29%).

The intra-rater reliability coefficients for DOS within medical admissions present similar findings. As indicated in Table III - 75, the specific non-acute reliability coefficient for DOS within medical admissions is higher for the AEP. The specific non-acute reliability coefficient for medical DOS for the AEP is 86% and 47% for the SMI.

CONCLUSIONS

Since this research indicates that the SMI is not reliable*, validity analysis of the SMI will not be performed. A necessary, but not sufficient, prerequisite for a valid criteria-set is that it be reliable. The SMI criteria is not reliable, therefore, it can not be valid.

---

* See section on reconsideration of the SMI reliability using nurse reviewers from a Professional Review Organization (PRO), Attachment E. Since there is limited evidence in support of the SMI reliability (using Rater "A") and because the findings of this research are substantially different from those reported by SysteMetrics, additional reliability trials with an equal distribution of reviews per reviewer and a larger number of reviews (N = 175) was performed (Attachment E).

While the AEP has been shown to be reliable, its validity has not been established. Before examining the validity of the AEP, we will examine a related issue: The use of overrides and the reliability of criteria.

AEP RELIABILITY AND OVERRIDES

The use of overrides is an important one. Overrides are used to change the final assessment due to clinical analysis and a host of other factors including those not considered by the objective criteria. Overrides, theoretically, provide clinical depth to an otherwise unidimensional set of criteria. Overrides can be used in one of two ways. An override can be used to find care to be non-acute even when "objective" criteria are met, or to assess care as acute when "objective" criteria are not met.

Theoretically, overrides can adjust for differences in community standards or the availability of sub-acute care. Overrides may also be used when objective criteria are inadequate in capturing the complexity of illness and medical treatment. Methodologically, the use of overrides allows reviewers to use clinical knowledge to "alter" outcomes rather than "fudge" or exaggerate the objective data. To the extent that overrides are the rule rather than the exception, the criteria-set verges on being subjective rather than purely objective and criteria-based. In addition, the use of overrides may affect reliability coefficients by reducing rater agreement due to the subjective nature of the use of overrides or because overrides may be used on more complex medical cases.

A more refined analysis of reliability with and without overrides will clarify the effect of overrides on reliability. To capture the purely objective

performance of these criteria, the following analysis is based on cases in which the override option was not employed. If an override was used, the case was dropped from the analysis. Thus, the reliability coefficients in this analysis represent the reliability without overrides. These probably represent less complex cases to evaluate.

As illustrated in Table III - 78, when the cases in which overrides are used are eliminated from analysis, the specific non-acute reliability for admission review, using the AEP, increases to 89% (kappa statistic = .78, $p < .01$), while the specific non-acute reliability for the SMI remains low at 14%. These data indicate that the use of overrides has minimal effect on the reliability of the SMI. However, for AEP admission reviews, the cases in which overrides are used tend to decrease the specific non-acute reliability of the AEP. This suggests that either the override use with the AEP instrument is less reliable than the objective criteria, and/or the cases on which overrides are used are more complex which results in lower reliability coefficients.

Similarly, the specific non-acute reliability for day review, using the AEP, increases to approximately 87%, while the specific non-acute reliability coefficient in cases in which the override option was not employed for the SMI, is approximately 35% (see Table III - 81). The use of overrides does not appear to affect the reliability of the SMI to any major extent.

These data provide two conclusions: (1) When cases in which the override options are employed are deleted from the data set, the reliability of the SMI remains low, and (2) The reliability of the AEP criteria improves. The cases in which overrides are used are more complex and/or the overrides themselves

are used less reliably than the more objective AEP criteria.

These results are strengthened by analysis of intra-rater reliability for both admission and days for the AEP and the SMI with and without overrides. As indicated in Table III - 84, the intra-rater reliability for the AEP admission reviews reaches approximately 90%, while the intra-rater reliability for admission reviews employing the SMI is approximately 35%. The intra-rater reliability for day review, in cases where the override is not employed for the AEP is 92% and 47% for the SMI (see Table III - 87).

These data support the conclusions that the AEP is a reliable criteria for identifying non-acute care, while the SMI is either unreliable or less reliable than the AEP. The difference in reliability does not appear to be due to the use of overrides. Rather, the difference appears to be the result of differential application between reviewers. (For an analysis of the expected weighted reliability for both the AEP and the SMI, see Attachment D.) We now turn to an analysis of the validity of the AEP.

TABLE III - 2

COMPARISON OF THE RELIABILITY OF THE SMI AND AEP
AS REPORTED BY SYSTEMETRICS (1) AND GERTMAN AND RESTUCCIA (2)

		AEP	SMI
Overall Agreement	Admission Review	N/A (3)	97.6%
	(n)		(124)
	Kappa Statistic		(K=75.7%, p < .001)
	Day Review	91.8 - 94.3%	
	(n)	(182 - 193)	
	Kappa Statistic	(p < .0001)	
Uncorrected Scores	Days Review	84 - 91%	98.7%
	(n)	(182 - 193)	(1,245)
	Kappa Statistic	(p < .01)	(K=79.8%, p < .001)
Specific Agreement	Admissions	N/A	67.35%
	(n)		(124)
	Days	73.1 - 79.3%	
	(n)	(182 - 193)	
Uncorrected Scores	Days	58 - 72%	62.5%
	(n)	(182 - 193)	(1,245)

(1) SysteMetrics. The valid and reliable measurement of non-acute hospital utilization in a nationally representative sample. Final report, HCFA Contract No. 500-80-0053, February 25, 1984.

(2) Gertman, P.M. and Restuccia, J.D. The Appropriateness Evaluation Protocol: A technique for assessing unnecessary days of hospital care. Medical Care, 19 (8), 855-871.

(3) Not available.

TABLE III - 3

OVERALL[1] AND SPECIFIC[2] INTER-RATER RELIABILITY

AS REPORTED BY SYSTEMETRICS[3]:

ADMISSION MODEL

		Rater #1		
		Acute	Non-Acute	Total
Rater #2	Acute	116 (93.54%)	1 (.80%)	117 (94.35%)
	Non-Acute	2 (1.61%)	5 (4.03%)	7 (5.64%)
	Total	118 (95.15%)	6 (4.83%)	124 (100%)

(1) Overall reliability (agreement) = 97.58% (112/124) (kappa = .76, $p < .001$).

(2) Specific reliability (agreement = 62.5%. (5/8)

(3) Adapted from SysteMetrics, The valid and reliable measurement of non-acute hospital utilization in a nationally representative sample. Deliverable No. 6, Final Report, HCFA contract no.: 500-80-0053, 1984 (pg. IV 32 through IV 33).

TABLE III - 4

OVERALL (1) AND SPECIFIC (2) INTER-RATER RELIABILITY

AS REPORTED BY SYSTEMETRICS (3):

DAY MODEL

		Rater #1		
		Acute	Non-Acute	Total
Rater #2	Acute	1,196 (96.06%)	4 (.32%)	1,200 (96.38%)
	Non-Acute	12 (.96%)	33 (2.65%)	45 (3.61%)
	Total	1,208 (97.03%)	37 (2.97%)	1,245 (100%)

(1) Overall reliability (agreement) = 98.71% (kappa = .80, $p < .001$).

(2) Specific reliability (agreement) = .67, 33/49

(3) Adapted from SysteMetrics (pg. IV 32 - IV 33).

TABLE III - 5

COMPARISON OF THE ADMISSION AGREEMENT/DISAGREEMENT MODEL
(NUMBER OF HOSPITALS INCLUDED IN RELIABILITY SAMPLE: 13)

Admission Review

SMI \ AEP	Acute	Non-Acute	Total
Acute	17 (17.17%)	68 (68.69%)	85 (85.86%)
Non-Acute	3 (3.03%)	11 (11.11%)	14 (14.14%)
Total	20 (20.20%)	79 (79.79%)	99 (100.00%)

COMPARISON OF THE DAY AGREEMENT/DISAGREEMENT MODEL

Day Reviews

SMI \ AEP	Acute	Non-Acute	Total
Acute	86 (23.69%)	168 (46.28%)	254 (69.97%)
Non-Acute	9 (2.48%)	100 (27.55%)	109 (30.03%)
Total	95 (26.17%)	268 (73.83%)	363 (100.00%)

TABLE III - 6

HYPOTHETICAL RELATIONSHIP BETWEEN INTER-RATER AND INTRA-RATER SPECIFIC NON-ACUTE RELIABILITY (1)

		SPECIFIC NON-ACUTE INTER-RATER RELIABILITY		
		Low Reliability (< 49%)	Moderate Reliability (50% - 74%)	High Reliability (> 75%)
INTRA-RATER RELIABILITY	Low Reliability	Inconsistent Raters Inconsistent Instrument (Worst Case Scenario)	Inconsistent Raters Probably Inconsistent Instrument	Inconsistent Raters Highly Consistent Instrument (Highly Unlikely Scenario)
	Moderate Reliability	Moderately Consistent Raters Inconsistent Instrument	Moderately Consistent Raters Moderately Consistent Instrument	Moderately Consistent Raters Highly Consistent Instrument
	High Reliability	Highly Consistent Raters Inconsistent Instrument (Highly Unlikely Scenario)	Highly Consistent Raters Moderately Consistent Instrument	Highly Consistent Raters Highly Consistent Instrument (Best Case Scenario)

(1) While absolute reliability values have not been established and indeed may vary, depending on the use of the criteria being evaluated for the purpose of identifying non-acute hospitalization, we consider reliability coefficients in excess of 75% to be outstanding; reliability coefficients between 50% and 74% to be moderate, and below 49% to be poor.

TABLE III - 7

DISTRIBUTION OF RANDOMLY SAMPLED CASES BY

INSTRUMENT AND OUTCOME

Admission Model

		AEP		
		Acute	Non-Acute	Total
SMI	Acute	758 (64.62%)	336 (28.64%)	1,094 (93.26%)
	Non-Acute	15 (1.29%)	64 (5.45%)	79 (6.74%)
	Total	773 (65.91%)	400 (34.00%)	1,173 (100.00%)

TABLE III - 8

DISTRIBUTION OF RANDOMLY SAMPLED CASES BY

INSTRUMENT AND OUTCOME

Day Model

		AEP		
		Acute	Non-Acute	Total
SMI	Acute	2,845 (55.33%)	1,408 (27.38%)	4,253 (82.71%)
	Non-Acute	125 (2.43%)	764 (7.64%)	889 (17.29%)
	Total	2,970 (57.76%)	2,172 (42.24%)	5,142 (100.00%)

Validity

VALIDITY

Procedures

Each of the cases sampled for reliability and validity (N = 172), described previously, were reviewed by twenty two (22) physicians. Eleven physicians (seven internists and four surgeons) were selected from a Michigan based, staff model, predominantly fee-for-service, large bedded, high quality teaching hospital located in the same community from which the sample cases and hospitals were selected (Henry Ford Hospital, Detroit, Michigan) and eleven physicians (seven internists and four surgeons) selected from a west coast, high quality, well respected, group model, health maintenance organization (Kaiser Permanente Medical Group, Oakland, California). The Michigan based physicians' assessments reflected general community medical standards, while the HMO based physicians assessed the appropriateness of care based on HMO practice and program standards. (Kaiser Permanente Medical Group of Northern California was chosen because they have demonstrated, in the published research literature, the lowest rates of hospitalization per 1,000 members of

any HMO in the nation.) While it could be argued that it is inappropriate for California based physicians to evaluate care performed in Michigan settings, the idea was to: (1) Assess whether any differences exist within and between Michigan based FFS physicians and California based HMO physicians, and (2) Determine the upper (HMO physicians) and lower limits (FFS physicians) of validity and estimates of non-acute care.

Each of the 172 cases sampled for validity analysis was reviewed by three FFS and three HMO physicians. Internists reviewed patient cases admitted for medical reasons, while surgeons reviewed cases admitted for elective surgery or cases in which a major surgical procedure was performed. Medical admissions for invasive diagnostic or treatment procedures (e.g., bronchoscopy) or for medical admission workup that resulted in surgery were also reviewed, where appropriate, by the surgeons. No attempt was made to match physician specialty with the diagnosis of the cases they reviewed. All physician assessments were independent and the cases assigned to them were random. Physicians were instructed to evaluate the appropriateness of each admission and assigned days of care review based on their independent clinical review. All physicians conducted blind reviews, (i.e., physicians did not know whether the care had been deemed to be acute or non-acute by the instruments).

Confirmation of Acute/Non-Acute Care

Confirmation or validation of the AEP reviews is not a straightforward matter. Physicians do not necessarily agree on whether an instance of care is acute or not. Since the leniency or stringency of the validation rule is a matter of policy rather than research, the validation or non-validation of the AEP is presented based on the extent of physician agreement. The most stringent decision rule for validating an AEP finding of non-acute care requires 100% agreement by physicians that care is non-acute. Conversely, the most lenient confirmation rule requires at least one physician (one out of three) to corroborate non-acute care. For purposes of this project it is argued that there is some support for the AEP when at least one physician agrees that care could have been performed in other than an acute-care setting and the validation of the AEP cannot be discounted. On the other hand, in a "real life" review (e.g., pre-certification and/or concurrent utilization review or retrospective appeals process), the majority rule (two out of three) would typically prevail. The validity of the AEP can be assessed on minimum (lenient), medium (moderate) or maximum (stringent) standards.

Summary of the Validity Results

The data will show that the validity of the AEP is limited. The tendency is for the AEP to find care non-acute while physicians find that same care to be acute. For conceptual and substantive reasons, we divide the validity assessment into four subsections: Medical Admissions, Days of Stay Within Medical Admissions, Surgical Admissions and Days of Stay Within Surgical Admissions. We find that the validity of the AEP in assessing the necessity of medical admissions ranges from 15% to 71%. The validity, depending on the

stringency of the validation rule, for days of stay within medical admissions ranges from 26% to 71%. The validity of the SAEP in assessing the necessity of surgical admissions ranges from 7% to 47% while the validity of the AEP in assessing days of care within surgical admissions ranges from 8% to 44%.

In general, these data demonstrate that the AEP is valid for medical admissions and medical days of stay only under the most lenient validation rule. Under the moderate rule, the AEP is validated for medical admissions only 36% and is validated for medical days of stay 54%. All three physicians agree with the AEP that an admission is non-acute only 15%. Similarly, under the most conservative validation rule, the days of stay assessment of the AEP is validated by three physicians 26%. The data indicate that the direction of disagreement between the AEP and physician reviewers is consistently in the same direction: **The AEP tends to find care non-acute when physicians tend to find that same care to be acute. This finding holds whether cases are validated by FFS or HMO physicians.**

Data is presented which suggests that the SAEP (Surgical Appropriateness Evaluation Protocol) is not a valid instrument for assessing the appropriateness of admissions for elective surgery. In addition, there is a pronounced tendency for the days of stay within surgical admissions not to be validated. It should be noted that the SAEP findings are considered tentative because of the small number of surgical observations in the sample ($N = 24$). We now turn to a detailed analysis of the validity data. [The supporting tables for the Validity section (Tables III - 90 through III - 147) may be found in Appendix B, Book II.]

FFS PHYSICIAN VALIDITY ASSESSMENT

Medical Admissions

As indicated in Table III - 90, for medical admissions (n = 148) the specific non-acute validity coefficient is, under the most lenient validation rule, 71% ($K = .43$, $p < .01$). While this level of validation is acceptable, two points must be made. First, once the stringency of the validation rule is increased requiring at least 2 of the 3 physicians to agree with the AEP, the specific non-acute validity falls to 36% ($K = .16$, $p < .01$). Secondly, once the validation rule requires 3 out of 3 physicians to validate a finding of non-acute, the specific non-acute validity coefficient of the AEP drops to 15% ($K = .07$, $p < .01$). These latter two validity coefficients are not considered sufficiently powerful to substantiate the AEP validity. Further, as indicated in Tables III - 91, III - 92 and III - 93, the directionality of the disagreement between the AEP and physicians is consistent. The AEP tends to find care non-acute while the physicians tend to disagree and find that same care to be acute. As indicated in Table III - 91, of the 117 instances in which the AEP assessed the admissions to be non-acute, FFS physicians agreed with the instrument assessment 87 times; however, of the 36 instances in which the physicians and the AEP instrument were not in agreement, physicians found the admission to be acute 30 of the 36 times. This pattern of disagreement between the AEP and the physician assessment was found under the moderate and conservative conditions as well.

These findings lend only minimal support to the validity of the AEP. The data also clearly indicate that fee-for-service physicians consistently find that the AEP makes false-positive identifications of non-acute medical admissions. These data can be interpreted in one of two ways. Either the AEP is not a valid instrument and finds only limited support among physicians with assessments of non-acute care because the AEP criteria for assessing the necessity of admissions is overly and inappropriately stringent, or the AEP is a valid instrument and the physicians' assessments are excessively conservative. While the latter interpretation is possible, great pains were taken to recruit first rate quality physicians representing the community standards for assessing the appropriateness of hospitalization. Secondly, the issue of the physician decision makers will be more fully addressed in the section which includes a validation assessment of the AEP by physicians employed in a large, well known, highly respected, group model, HMO organization (Kaiser Permanente Medical Group of northern California).

Medical Days Of Care

Having just demonstrated the validity of the AEP for assessing the necessity of medical admissions is limited, we turn to the validation of the AEP in assessing the necessity of days of care within medical admissions. The data will show that the AEP is more valid in assessing days of care within medical admissions than in assessing the necessity of the medical admission itself. The data provide support for the validation of the AEP DOS criteria under the lenient and moderate conditions.

As indicated in Table III - 94, the specific non-acute validity of the AEP in assessing days of care ranges from a low of 26% to a high of 71%. The non-acute validity coefficient under the lenient rule is 71% ($K = .43$, $p < .01$), under the moderate rule the AEP is validated 54% ($K = .30$, $p < .01$), and under the most conservative rule the AEP is validated 26% ($K = .12$, $p < .01$).

These data provide support for the validity of the AEP in assessing non-acute days of care under the lenient and to some extent under the moderate rule. However, the AEP overstates the degree of non-acute days of care from 30% to 75% depending on the FFS physician agreement rule implemented. **These false-positive findings indicate the pronounced tendency of the AEP to identify days of care to be non-acute while FFS physician reviewers do not consider that same care to be non-acute.**

Insert Table III - 94 About Here

Tables III - 95, III - 96 and III - 97 indicate the direction of the disagreement between the AEP and physicians is consistent except for the lenient rule. Under the lenient rule, of the 516 days of stay found by the AEP to be non-acute, physicians found 92 days to be acute. Similarly, under the moderate rule, of the 516 instances in which the AEP found days of care to be non-acute, the physicians found 212 days to be acute. Finally, under the most conservative rule, of the 516 days of care found to be non-acute by the AEP, physicians found 379 days of care to be acute.

Insert Tables III - 95, III - 96 and III - 97 About Here

These data provide limited support for the validity of the AEP in assessing the necessity of individual days of stay under the lenient and moderate rules. We now turn to an assesssment of the newly developed Surgical Appropriateness Evaluation Protocol (SAEP).

Surgical AEP: Elective Surgical Admissions

Before analyzing the SAEP validation data, it is important to note that the results should be considered preliminary because of the small number of surgical admissions in the validation sample. There are only 24 elective admissions in the surgical validation sample. Of these 24 admissions, the SAEP found 15 admissions to be non-acute. To provide more definitive results, it would require 30 to 50 instances in which the SAEP finds care (admissions) to be non-acute for firm conclusions to be drawn. Nevertheless, if the SAEP is a reliable and valid instrument, it is expected that the preliminary data will provide support for the SAEP. The data, however, do not support the contention that the SAEP accurately and appropriately identifies non-acute admissions for elective surgery. We now present the data supporting these findings.

The data indicate that FFS physicians do not validate the SAEP criteria for assessing the appropriateness of elective surgical admissions (see Table III - 98).

Insert Table III - 98 About Here

The specific non-acute reliability coefficient is 47.06% under the lenient rule and 26.67% under the moderate rule. The data, on a limited number of cases (N = 24), suggest that FFS surgeons do not agree with the assessment of the SAEP on elective surgical admissions. Combining this finding with the moderate reliability of the SAEP (60%), we find a criteria-set that is not recommended for use without further refinement and development.

DOS Within Elective Surgical Admissions

The data indicate that the ability of the AEP DOS criteria-set is minimally (under the lenient rule) valid for assessing the appropriateness of continued surgical stay. As indicated in Table III - 102, the kappa statistic is significantly better than chance association for the lenient validity rule (kappa = .22, $p < .01$). However, the kappa is only marginally significant and the specific non-acute reliability coefficient is 44.12%. Once the validity rule is tightened to require agreement between at least two (2) physicians, the validity coefficient for specific non-acute assessments drops to 24.53%.

Insert Table III - 102 About Here

These data are remarkable for several reasons. The validity assessment for surgical admissions (47.06%) under the lenient rule, is substantially lower than the lenient validity coefficient for AEP medical admissions (70.73%). Similarly, the validity coefficient under the lenient rule, for DOS within surgical admissions (44.12%) is substantially lower than the validity coeffi-

cient for DOS within medical admissions (71.02%), under the lenient rule. Our original expectations were that indications for surgery and continued post surgical recuperation would be easier to assess than medical admissions and continued DOS. There is some indication that physicians may agree with the AEP medical criteria but do not agree with the surgical criteria or with the medical criteria for continued DOS within surgical admissions. This is especially interesting since the criteria for assessing both medical and surgical DOS are identical.

Finally, it is important to note that in all medical and surgical admissions and DOS reviews, when physicians disagree with the AEP/SAEP criteria, physicians, even under the lenient rule, generally find care to be acute, while the AEP/SAEP criteria generally find care non-acute. The AEP/SAEP admission and DOS criteria, when there is disagreement with physicians, **overestimates** non-acute care according to FFS physicians. Thus, AEP/SAEP rates of non-acute care in studies assessing rates of non-acute care or in pre-admission certification reviews, according to the FFS physicians in this study, overestimate the extent of non-acute care. This may result in inappropriate admission denials and continued stay certification in pre-determination utilization review and other cost-containment programs. We now turn to an assessment of the criteria by physicians employed in an HMO setting.

HMO PHYSICIAN VALIDITY ASSESSMENT

AEP Medical Admissions Criteria

The data provide support for the validity of the AEP medical admission criteria by HMO physicians. The specific non-acute validity coefficient for medical admissions by HMO physicians is 77.24% under the lenient rule, and 61.67% under the moderate rule (see Table III - 106).

Insert Table III - 106 About Here

AEP DOS Within Medical Admissions

Similarly, the HMO physicians, in general, validate the AEP DOS criteria under both the lenient (specific non-acute validity = 73.91%) and moderate rules (specific non-acute validity = 65.87%, see Table III - 110). While HMO physicians tend to validate the AEP medical admission and DOS criteria, when disagreements between physicians and the criteria-set are found, the HMO physicians, like the FFS physicians, in general, find care to be acute while the AEP reports care as non-acute. **Thus, even when validated by HMO physicians, the data indicate that the AEP criteria overestimate non-acute care** (see Tables III - 111, III - 112 and III - 113).

Insert Tables III - 110, 111, 112 and 113 About Here

SAEP Surgical Admissions and DOS Within Surgical Admissions

The HMO physicians, in general, do not validate the SAEP criteria for elective surgery admissions, but do validate the DOS within elective surgery criteria. As indicated in Table III - 114, under the lenient rule for elective surgery admissions, while specific non-acute validity reaches 60.87%, the kappa statistic of .05 is not significantly better than zero (chance association). Interestingly, the patterns of agreement/disagreement between HMO physicians and the SAEP admission criteria is entirely in the direction of the physicians finding significantly more non-acute care than the SAEP criteria. HMO physicians' standards, in contrast to FFS physicians' standards, do not find the SAEP valid because the SAEP tends to inappropriately underestimate non-acute surgical admission (see Tables III - 115, III - 116 and III - 117).

Insert Tables III - 114, 115, 116 and 117 About Here

The HMO physicians, in contrast to the FFS physicians, tend to validate the AEP DOS criteria for elective surgery days. As demonstrated in Table III - 118, the specific non-acute validity of days within surgical admissions is 59.15% and 48.33% under the lenient and moderate rules, respectively. When HMO physicians disagree with the AEP DOS criteria for surgical admissions, the physicians indicate care to be acute while the criteria indicate care as non-acute. Thus, the AEP DOS criteria for surgical days also overestimate non-acute care.

Insert Table III - 118 About Here

CONCLUSIONS

FFS AND HMO VALIDATION*

The data indicate that the AEP medical admission criteria are validated by both HMO and FFS physicians under the lenient rule. The criteria are validated to a greater extent by HMO than FFS physicians, especially under the moderate rule. Similarly, FFS and HMO physicians validate the AEP DOS criteria to an equal extent (see Table III - 122). The SAEP criteria are not validated by either FFS or HMO physicians. The AEP criteria for DOS within elective surgical admissions are only marginally validated by FFS physicians, while HMO physicians validate the surgical DOS criteria to a greater extent than FFS physicians (see Table III - 123).

Insert Tables III - 122 and III - 123 About Here

These data indicate that the AEP/SAEP admission and DOS criteria, in general, overestimate non-acute care according to both FFS and HMO physicians. In addition, the data indicate that the AEP/SAEP criteria are more indicative of the hospitalization patterns practiced in an HMO as opposed to FFS settings. Previous research, as reported in Luft (1981), indicates that HMO physicians located in the western United States have the lowest rates of hospitalization per capita in comparison to physicians in other geographic or organizational

---

* While it is extremely interesting to compare the reliability of FFS and HMO physicians, similar to the comparison of the AEP and SMI reliability, this aspect of the technical report is currently being prepared as an article to be considered for publication (in progress).

settings. Thus, the validation of the AEP/SAEP by these HMO physicians should be considered as the theoretical maximum validation of the AEP criteria. The criteria's validity, in application, may best be represented by the FFS physicians, since they represent the norms of the community in which the care was provided. These physicians provide only limited validation of the AEP criteria and little validation of the SAEP and AEP DOS surgical criteria.

Finally, original data on the differences between FFS and HMO physicians admission and discharge practice patterns are provided. The probable reason for the dramatic differences in hospitalization rates per capita are the result of significantly lower non-acute admissions, especially surgical admissions, and to some extent DOS by HMO physicians in comparison to FFS physicians. Although the subject of another paper now in progress (Strumwasser & Paranjpe), another major difference between HMO and FFS physicians is the homogeneity of physician practice patterns. In FFS settings, when one of the physicians find care to be non-acute, there is not a significant likelihood that one of the other two physicians will also find care to be non-acute. On the other hand, the homogeneity of HMO physician practice patterns is significantly greater than in FFS settings. When one HMO physician finds medical or surgical admissions and/or DOS to be non-acute, there is a significant likelihood that one of the other two remaining physicians will also find the care to be non-acute. Of course, the remaining question is whether there is any difference between FFS and HMO settings in the quality of the hospital care (Strumwasser, Paranjpe, Dmuchowski and McGinnis, in progress) rendered/not rendered (over hospitalization/under hospitalization).

We now turn our attention to an analysis of the estimates of non-acute care as adjusted by FFS and HMO physician validation under lenient, moderate and conservative decision rules.

Estimates Of Non-Acute Care*

---

* Portions of this section appear in an article by Strumwasser, I., Paranjpe, N.V., and Hall, H. "Determining Nonacute Hospital Stays". Business and Health, February, 1987.

ESTIMATES OF NON-ACUTE CARE

As financial resources become limited, controlling or managing health care costs becomes increasingly important. A constructive method of reducing costs without threatening the quality of health care is to reduce non-acute hospital use. Reduction of non-essential inpatient hospital services ensures that cost savings are not achieved by limiting access to necessary inpatient care. Resource management programs are the key to the success of managed health care delivery and cost containment programs.

This chapter presents estimates of non-acute hospitalization in a major metropolitan area and recommends the design of utilization review programs based on empirical research findings. Non-acute hospitalization refers to hospital care that could have been provided in other than an acute care, inpatient hospital setting. The AEP/SAEP utilization review criteria used in this section assess the appropriateness of the **setting** or location of services and **not the appropriateness or quality of the care provided.**

One example of a non-acute medical admission is a hospital admission for low back pain, with orders for oral pain medication and progress notes indicating the patient is ambulating. In this case, the admission and most, or all, of the days of hospital care would probably be considered non-acute, since the indicated non-surgical course of therapy is typically bed rest and pain control, which can safely and appropriately be provided in other than an acute care setting.

Acute medical admissions with non-acute days frequently reflect institutional inefficiencies such as unnecessary delays in the results of lab tests, or problems with discharge planning. Consider the case of an appropriately admitted, newly diagnosed diabetic with fasting blood sugars of 500 mg. per dl., ketones in urine and on daily insulin dose adjustments. Once the patient's insulin dose is regulated and there is no danger of diabetic coma, the patient is ready for discharge. However, discharge is delayed for diabetic education (diet and self-administration of insulin), which could have been conducted earlier during the hospital stay or on an outpatient basis.

A general example of a non-acute medical admission with acute days of care might be an inpatient admission for diagnostic testing which could have appropriately and safely been performed on an outpatient basis. If the results of this testing indicate the need for inpatient surgery, which is then performed, the days subsequent to the surgery would probably be considered acute until utilization review indicates that the patient is ready for discharge. Thus, it is possible to have acute days within an otherwise non-acute hospital admission.

A final example of a non-acute admission for surgery includes hospital admissions for procedures which can safely and appropriately be performed on an outpatient basis (e.g., cataract surgery on an otherwise healthy individual), or for an elective surgery admission admitted several days prior to routine elective surgery.

As is evident from these examples, criteria which evaluate the appropriateness of the setting in which care is provided do not necessarily reflect the need for care nor the quality of the care performed. In some instances, no acute care is required. In other instances, sub-acute care may be appropriate. Non-acute hospital care may also reflect unnecessary hospital care and costs (e.g., when discharge is appropriate but unnecessarily delayed).

A commitment to high quality, cost-conscious and accessible health care requires the use of reliable and valid criteria for identifying the appropriateness of treatment in an inpatient hospital setting. It is critical that users of these protocols know what these criteria-sets can and cannot do, how well they perform their function, whether or in what capacity they should be used, and what they tell us about non-acute hospitalization and health care expenditures.

The Appropriateness Evaluation Protocol (AEP) and the Surgical Appropriateness Evaluation Protocol (SAEP), objective criteria sets used to assess the appropriateness of the setting in which acute hospital care is provided, were applied to 1,266 randomly selected cases drawn from the population of all 1983 admissions (excluding pediatric, psychiatric and obstetric) in 21 randomly sampled acute-care hospitals located in a six county area in southeast Michigan. Hospital selection was based on a multi-probability sample stratified by teaching status, bed-size and occupancy rate. The probability of hospital selection was proportional to the number of admissions to the facility.

The 21 hospitals in the study represent 317,109 (45%) of all 1983 admissions (712,048) to the 80 acute-care facilities in southeast Michigan. These 80 hospitals account for approximately 70% of all admissions to all acute-care facilities in the State of Michigan. Inpatient episodes (cases and days of stay) chosen for analysis were based on a random sample. Thus, estimates of non-acute care among hospitals are representative of all admissions to all hospitals in the study area during 1983. Finally, within hospital case selection was based on the number of admissions to a given hospital by payor source (BCBSM, Medicare and Medicaid).

To provide an analysis of the extent of non-acute care, the reliability and validity of the utilization review criteria needed to be established. The results of the reliability trials indicate that the AEP/SAEP review criteria are highly reliable in assessing the appropriateness of both admissions and days of stay (DOS) to inpatient hospital settings. The agreement between raters on findings of non-acute is 75% for admissions and 80% for days of stay. The reliability of the elective surgery admissions criteria is more modest at a 60% agreement rate.

The criteria for use in this section were validated by salaried physicians employed by a large, teaching, staff model, fee-for-service (FFS) hospital located in southeast Michigan. Three physicians reviewed each case. The instrument/RN non-acute finding was considered corroborated in this chapter if any two of the three physicians reviewing the case agreed that care was non-acute. Validation of the criteria by one out of three FFS and by HMO physicians is also provided (see Summary Tables III - 148 and III - 149). The data indicate that the criteria tend to overestimate the extent of non-acute care.

Physicians agree with admissions identified by the criteria to be non-acute about 36% of the time. The day of stay criteria were supported by independent physician assessment at a rate of 55%. The surgical admission criteria were validated at a rate of about 27%.

The important point is that the criteria tend to report a significant number of false-positive identifications of non-acute care. When physicians disagree with the criteria, they almost always identify the care as acute, while the criteria identifies the care as non-acute. This finding has important implications for designing a comprehensive utilization review program. For example, the AEP should never be used to deny payment or for pre-authorization of admissions to, or continued days of stay in, acute care hospitals without physician confirmation because of the tendency of the criteria toward false-positive identification of non-acute care. Inappropriately used, the criteria could have a profound negative effect on the quality of care.

In conjunction with physician review, or adjusted for false-positive identifications, the criteria can appropriately be used to identify the non-acute use of hospital resources in utilization review studies, for diagnostic, physician and/or hospital resource monitoring and case-management, or for use in pre-admission authorization programs. The information generated by utilization review criteria can also be useful in developing intervention strategies and for evaluating the success of managed care and utilization review programs. Adjusted estimates of non-acute care (see Tables III - 148 and III - 149, at conclusion of this chapter) are obtained by taking the rate of non-acute care determined by application of the criteria (see end of estimation chapter) and multiplying by the rate of fee-for-service physician validation

(see Table III - 150). The result is an empirically derived and valid rate of non-acute hospital care. Thus, estimates of non-acute care employing the standards of care and hospital practice patterns of FFS and HMO physicians under different rules may be obtained by selecting the desired organizational setting and decision rule and multiplying the validity by the unadjusted rates of non-acute care obtained by direct application of the AEP/SAEP (see estimation Appendix, Book III).

Insert Tables III - 148, 149 and 150 About Here

Approximately 12% of all medical admissions (see Figure III - 1) and 22% of all days of stay (see Figure III - 2) within medical (non-surgical) admissions did not require inpatient hospital level care. Significant savings could be generated by reducing or eliminating non-acute medical admissions and DOS. The extent of non-acute elective surgical admissions and DOS within elective surgery admissions appears to be minimal. Five (5) percent of surgical admissions (see Figure III - 1) and 7% of DOS within elective surgical admissions (see Figure III - 2) are found to be non-acute. Utilization review and managed care programs may effectively contain costs by focusing on medical rather than surgical admissions (see Figure III - 1).

Insert Figures III - 1 and III - 2 About Here

When these data were examined for the extent of non-acute days within acute and non-acute admissions, it was found that 16% of the days within acute medical admissions were non-acute, while 43% of the days within non-acute

medical admissions were non-acute. Similarly, approximately 5% of the days within acute elective surgical admissions were non-acute, while 20% of the days within non-acute surgical admissions were non-acute (see Figure III - 3).

Insert Figure III - 3 About Here

These data suggest that the best and most efficient way to eliminate non-acute hospital care is to reduce non-acute medical admissions. They clearly underscore the need for a quality pre-admission or pre-authorization program if health care costs are to be contained without sacrificing the quality of health care.

These data suggest that prospective payment and DRG (diagnostic related groups) reimbursement will eliminate only a small portion of the unnecessary days of stay. Since DRGs pay the hospital a fixed amount per admission, they provide some incentive to reduce non-acute DOS, but do nothing to contain costs by eliminating non-acute hospitalization which is where most of the non-acute days occur. For DRGs to be effective in reducing unnecessary hospital days, prospective payment should be implemented within the context of a pre-determination program. Such a program would eliminate the need for continued stay review (except for an outlier policy), as long as reimbursement rates are appropriately set. However, it is critical to remember, because of the large number of false-positive findings of non-acute care, that all denials of pre-authorization, using the AEP/SAEP, require review by qualified and knowledgeable physicians before denial of admission or payment.

These data provide empirically derived estimates of hospital specific rates of non-acute care. Policymakers and program developers may use this type of data to make decisions for setting acceptable targets and goals for hospital efficiency. Comparative rates of non-acute care tell us which hospitals have excessive rates of non-acute hospitalization as well as indicating which hospitals are most efficient and cost-conscious. Hospitals with excessive rates of non-acute care can be targeted for close utilization review inspections and other sanctions. Inefficient hospitals may also be considered for exclusion from preferred provider organizations (PPOs) and HMO (Health Maintenance Organizations) networks. Finally, with the involvement and cooperation of physicians and hospitals, and with appropriate reimbursement incentives, resource management programs and alternative health care delivery models such as PPOs, it may be reasonable to expect/require hospitals to attain, over time, and with proper monitoring and incentives, the efficiency rate of the most efficient hospitals within a given community.

These data indicate that important differences in rates of non-acute care exist among hospitals. Rates of non-acute hospitalization range from a low of 6% to a high of 25%, while the average rate of non-acute hospitalization over all hospitals is 12% (see Figure III - 4). Hospitals above the 12% non-acute care rate should be considered targets for special utilization review sanctions. Approximately 42% of the sample hospitals in the geographic area under examination have rates of non-acute care above the 12% level.

Insert Figure III - 4 About Here

Hospital group 1, representing about 29% of the sample, has the lowest rates of non-acute care ranging from 6% to 10%. These hospitals practice the most cost-conscious hospital care in the population. Group 2, representing about 33% of the hospitals, exhibits rates of non-acute care between 11% and 13%. While these rates may still be unacceptably high, they represent the average level of hospital and physician efficiency. It is perfectly reasonable to expect other hospitals, within the same community, to exhibit non-acute admitting practices similar to the community average.

On the other extreme, group 4 and 5 hospitals have extremely high absolute and comparative rates of non-acute care ranging from 18% to 25%. These hospitals represent approximately 19% of the population of the hospitals in the geographic area under study and the least efficient and cost-conscious hospitals in the community. Facilities exhibiting these high levels of non-acute hospitalization should be considered for special utilization review and other program sanctions. Their inclusion in PPO arrangements and open panel HMOs should be carefully considered. Group 3, representing about 19% of the population, exhibits rates of non-acute hospitalization between 14% to 16%. These hospitals should, with little effort, be able to bring their rates of non-acute hospitalization within reasonable and acceptable levels.

Utilization review data are very important in identifying target levels of non-acute care and hospital inefficiencies. The data also indicate that important differences exist in respective rates of non-acute medical care between teaching and non-teaching hospitals.

While the non-acute elective surgery rates of care are similar between teaching (6%) and non-teaching (5%) hospitals, teaching hospitals have a significantly lower rate of non-acute medical admissions (9%) than non-teaching (14%) hospitals. These data also indicate that teaching facilities which, on the average, may treat more severely ill patients, and are thus penalized under PPS, also have, on the average, less non-acute hospital care. These findings suggest that important, typically university affiliated, teaching facilities may face stressful times under the prospective payment system since they have less non-acute hospital care that can safely be eliminated (see Figure III - 5).

Insert Figure III -5 About Here

Finally, the data indicate that there is little difference in rates of non-acute care among payor sources. The rates of non-acute medical/elective surgical admissions are essentially equivalent among Medicare (12%/4%) and Medicaid (10%/5%) beneficiaries, and BCBSM (13%/6%) enrollees (see Figure III - 6).

Insert Figure III - 6 About Here

This research confirms the reliability of one method for assessing non-acute hospitalization (the AEP), while placing constraints on its appropriate use due to the high false-positive identification of non-acute hospital care. These data underscore the importance of physician review in all instances in

which hospital care is identified as non-acute. Independent physician assessment is especially important in pre-admission, pre-authorization managed care programs. We have demonstrated the usefulness of such criteria and data, when used appropriately, in identifying appropriate cases and hospitals for focused review activities and for the inclusion/exclusion of hospitals in alternative service delivery network (e.g., PPO) arrangements. Our research suggests that a major cost reduction can be obtained by instituting a pre-admission, pre-authorization review program in conjunction with a DRG type prospective payment system to limit the extent of non-acute days of stay. Pre-authorization programs will maximize their effectiveness by focusing on non-acute medical admissions, especially in non-teaching facilities.

Future research should focus on individual physician practice patterns within different hospital settings, as well as the change in levels of non-acute hospital care as a result of the federal government's prospective payment system (PPS) and as a result of pre-admission, pre-authorization programs. Such programs designed to reduce non-acute hospitalization and contain health care costs while maintaining access to quality health care, developed by leaders and innovators in cost containment, are vital to the health of the health care industry. [Strumwasser and Paranjpe, in progress, are currently conducting a study designed to measure the change in non-acute hospitalization between 1983 and 1986 (before and after the implementation of PPS).]

TABLE III - 148

GROSS ESTIMATES OF NON-ACUTE CARE MEDICAL ADMISSIONS

ADJUSTED FOR HMO AND FFS VALIDITY*

	VALIDITY RULE	HMO PHYSICIANS	FFS PHYSICIANS
ADMISSIONS	LIBERAL	25.98	23.79
ADMISSIONS	MODERATE	20.74	12.15
ADMISSIONS	CONSERVATIVE	6.50	4.89
DAYS OF CARE	LIBERAL	31.00	29.79
DAYS OF CARE	MODERATE	27.63	22.27
DAYS OF CARE	CONSERVATIVE	16.21	10.78

* In percent.

TABLE III - 149

GROSS ESTIMATES OF NON-ACUTE CARE SURGICAL ADMISSIONS

ADJUSTED FOR HMO AND FFS VALIDITY*

	VALIDITY RULE	HMO PHYSICIANS	FFS PHYSICIANS
ADMISSIONS	LIBERAL	12.45	9.63
ADMISSIONS	MODERATE	11.69	5.46
ADMISSIONS	CONSERVATIVE	11.51	1.36
DAYS OF CARE	LIBERAL	16.70	12.46
DAYS OF CARE	MODERATE	13.65	6.93
DAYS OF CARE	CONSERVATIVE	9.06	2.35

* In percent.

AEP/SAEP SPECIFIC NON-ACUTE VALIDATION COEFFICIENTS:

FFS AND HMO STANDARDS

FFS

	Medical Admissions	Surgical Admissions	Medical DOS	Surgical DOS
Rule 1 (Lenient)	70.73%*	47.06%	71.02%*	44.12%*
Rule 2 (Moderate)	36.13%*	26.67%	54.29%*	24.53%
Rule 3 (Conservative)	14.53%*	6.67%	25.70%*	8.33%

HMO

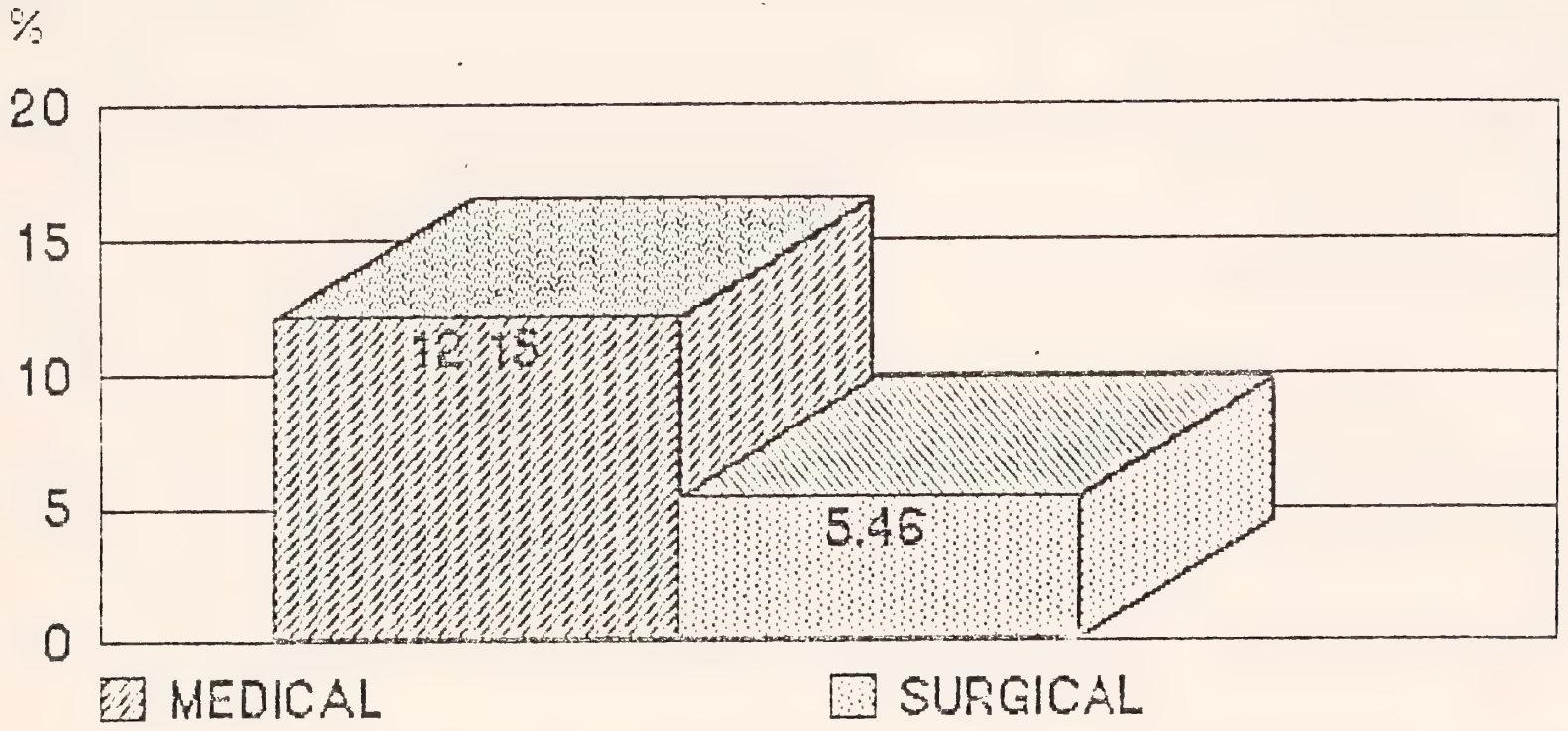
	Medical Admissions	Surgical Admissions	Medical DOS	Surgical DOS
Rule 1 (Lenient)	77.24%*	60.87%	73.91%*	59.15%*
Rule 2 (Moderate)	61.67%*	57.14%	65.87%*	48.38%*
Rule 3 (Conservative)	19.33%	56.25%*	38.66%*	32.08%*

* Kappa Statistic significant, p < .01.

(P7HMO)

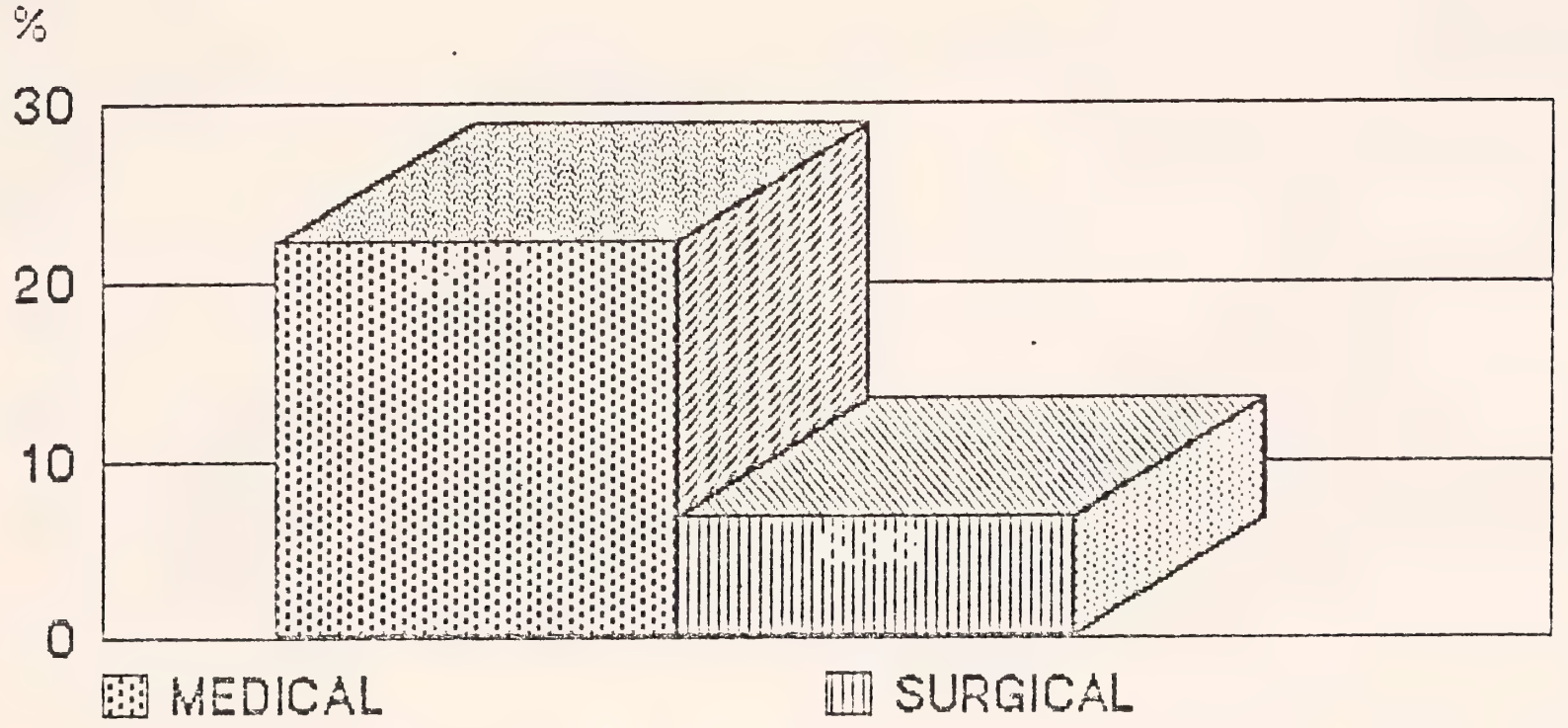
ESTIMATES OF NON-ACUTE CARE

ADMISSIONS



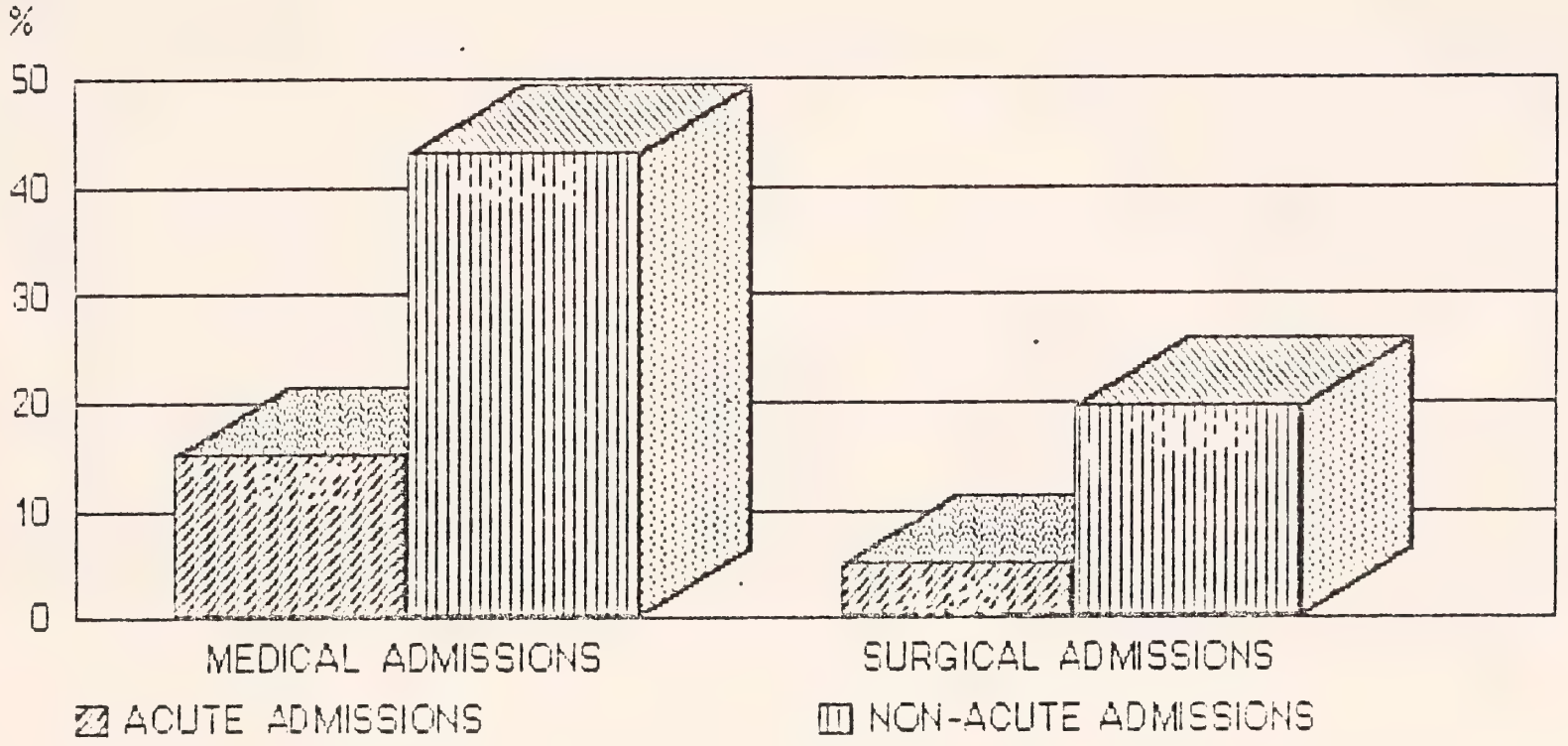
ESTIMATES OF NON-ACUTE CARE

DAYS OF CARE



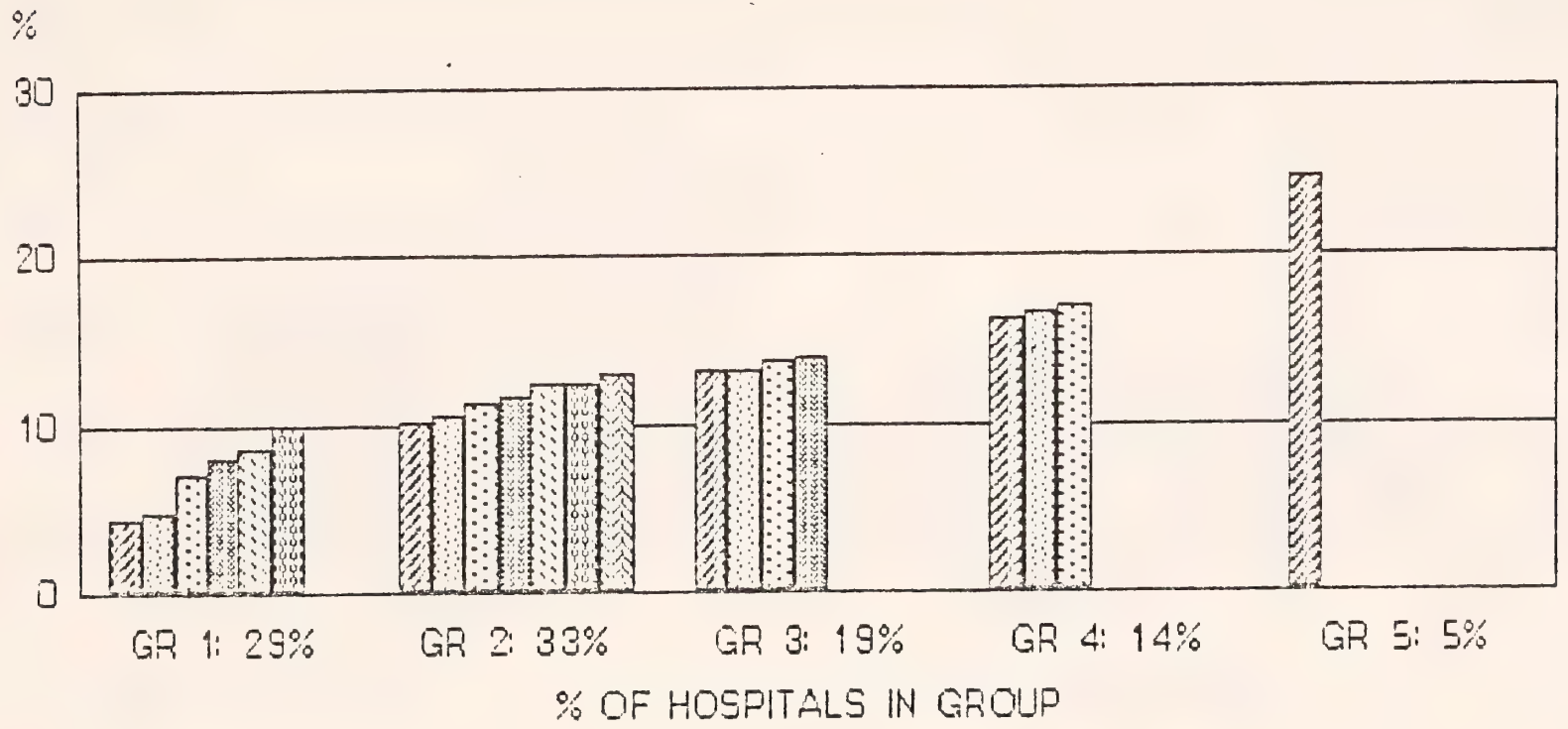
ESTIMATES OF NON-ACUTE DAYS

WITHIN ACUTE/NON-ACUTE ADMISSIONS



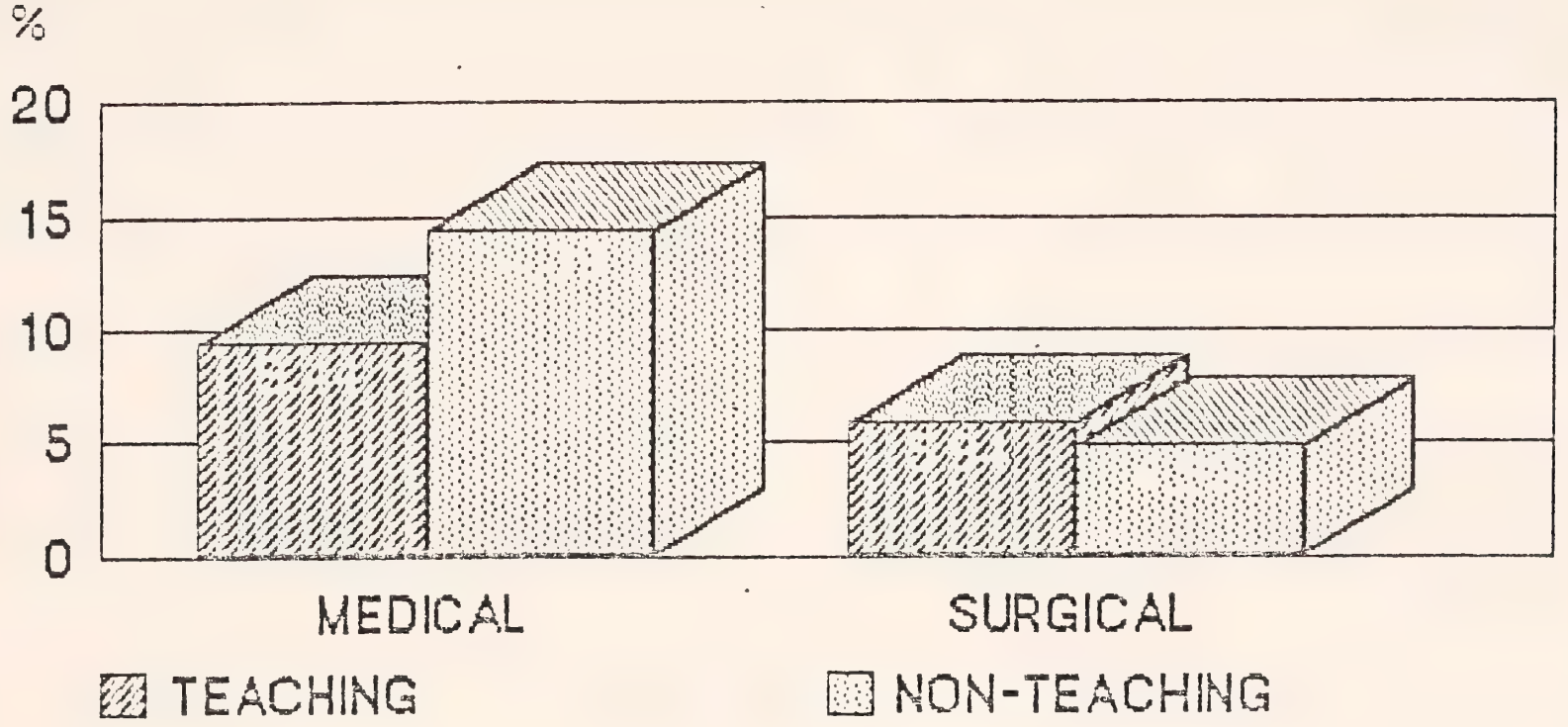
ESTIMATES OF NON-ACUTE CARE

MEDICAL ADMISSIONS BY HOSPITAL



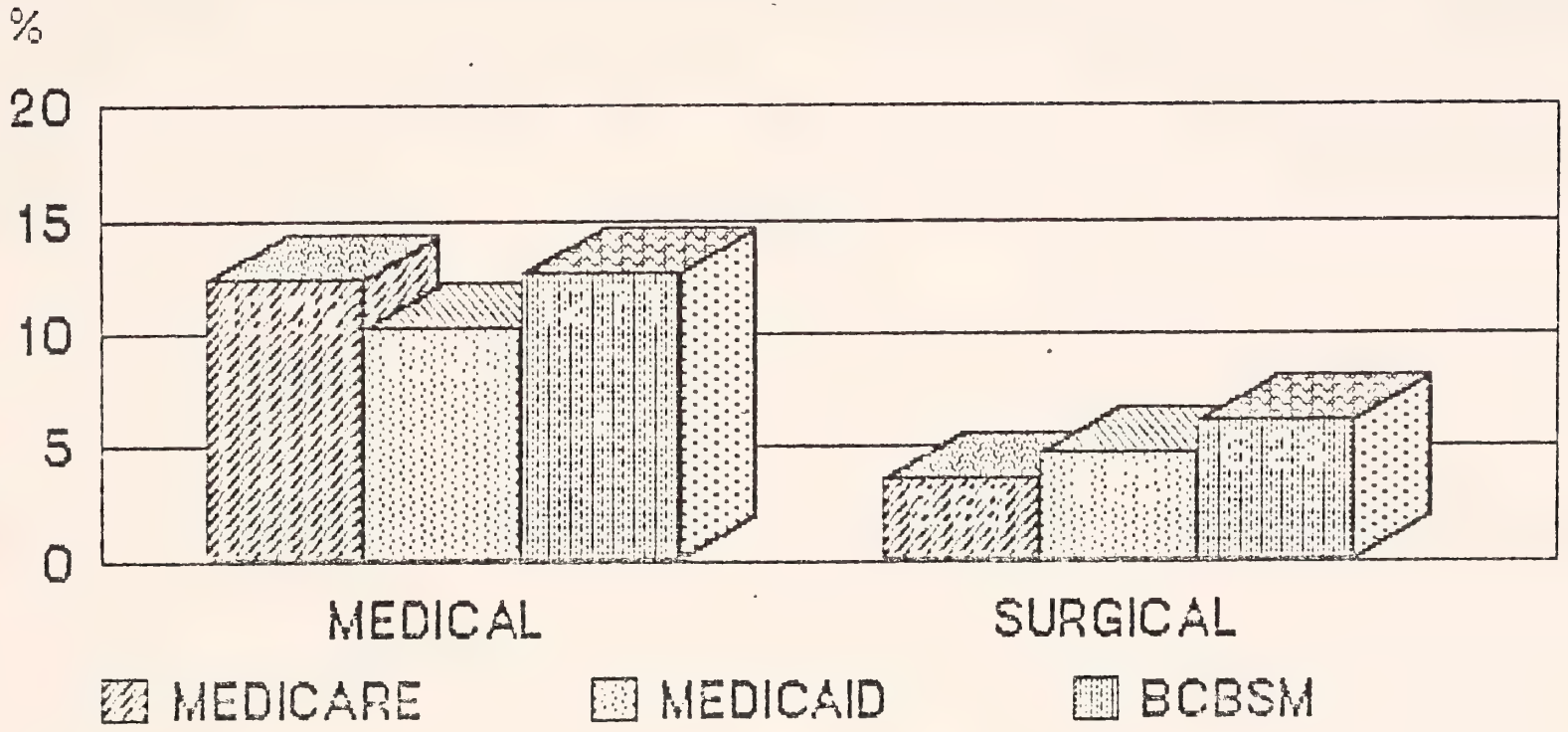
ESTIMATES OF NON-ACUTE CARE

ADMISSIONS BY TEACHING/NON-TEACHING STATUS



ESTIMATES OF NON-ACUTE CARE

ADMISSIONS BY PAYOR



CHAPTER IV

CONCLUSIONS

CONCLUSIONS

In this study, designed to assess both the reliability and validity of the Appropriateness Evaluation Protocol (and the Surgical Appropriateness Evaluation Protocol) and the Standardized Medreview Instrument, we have determined that the AEP is a reliable instrument for assessing the appropriateness of medical (non-elective surgical) admissions and medical days of stay to acute care facilities. The Surgical Appropriateness Evaluation Protocol, while being less reliable than the AEP, is found to be moderately reliable for evaluating the appropriateness of elective surgery admissions to acute care facilities. The data also indicate that the AEP criteria for evaluating the appropriateness of days of stay within elective surgery admissions are reliable.

The data presented in this research indicate that the Standardized Medreview Instrument is not a reliable criteria for evaluating the appropriateness of admissions and/or days of stay to acute care facilities when applied by RN's newly trained in its use. While the SMI is more reliable when applied by RNs with over two years experience with the criteria-set, it is still less reliable than application of the AEP by a group of RNs newly trained in use of the AEP (see Attachment E). Based on the relative and comparatively low reliability of the SMI, its continued use, without modification and improved reliability, especially within organizations without extensive experience in applying the SMI, is not recommended.

While the AEP and the SAEP criteria have been found to be reliable, the degree of validity depends on the stringency of the rule confirming a non-acute finding and whether or not physician validators are employed in a fee-for-service or an HMO setting. In general, regardless of the validation rule or organizational setting in which physicians are employed, the AEP appears to overestimate both non-acute admissions (medical and elective surgical) and days of stay. These false-positive identifications of non-acute care range from 29% to almost 85% as determined by fee-for-service physicians. A similar range of overestimation is corroborated by HMO physicians. Our best prudent estimation of the extent of false-positive identifications for the AEP/SAEP criteria is approximately 64%. This means that in more than half the cases in which care is identified as non-acute by the AEP, FFS physicians will not agree with the AEP assessment. In general, these findings hold for both fee-for-service and HMO physician validators (on the average HMO physicians, under a prudent rule, disagree with the AEP in 38% of the instances in which RNs applying the criteria find medical care to be non-acute).

These findings place specific limits and constraints on the use of the AEP criteria. In general, these criteria are used for two specific purposes. They are typically used in evaluative studies or in pre-admission certification and continued stay reviews, as well as retrospective payment review. The data suggest that findings provided by the AEP criteria should be adjusted depending on their specific use. For example, in studies which are directed toward an evaluation of the extent of non-acute hospital care, the non-acute AEP rates need to be adjusted downward, depending on the policy decision rule, to account for the false-positive overestimations of non-acute care. Thus,

for use in evaluative studies or for use in decisions regarding inclusion/exclusion of hospitals and providers in preferred provider arrangements, the data need to be adjusted appropriately.

If the AEP is used in pre-admission review, continued stay utilization or retrospective payment review, to prevent inappropriate denial of payment, and the potential negative effect on the quality of health care, **100% of all denials must be reviewed by physician UR consultants**. Given this caution, we find the AEP criteria acceptable for use as a UR instrument as long as special cautions against false-positive (incorrect) identifications of non-acute care are implemented.

Adoption of the moderate decision rule indicates that a substantial amount of hospital care appears to be non-acute. While the data represent non-acute care during the year 1983, no empirical data exist documenting the extent of non-acute hospital care during 1986. Substantial rates of non-acute care and excess hospital care costs exist and substantial savings can be realized by more efficient and cost conscious delivery of hospital level resources. What remains unclear is the change in patterns of non-acute care that have resulted as a result of the federal government's prospective payment system and from a host of cost containment activities that have flourished since 1983 (i.e., BCBSM's pre-admission, pre-certification and continued stay certification). We know very little about how these cost containment programs such as pre-admission, pre-certification review and other managed care programs have altered rates of non-acute care. For a complete discussion of this issue, and the need for continued research to answer these complex questions, the reader is referred to a research project in progress by Strumwasser and Paranjpe,

1986 [Estimates of Non-Acute Hospitalization: "Then (1983) and Now (1986)"].

While the current study evaluated the extent of reliability and validity of the AEP and the SMI, very little attention has been paid to a third utilization instrument known as the Intensity, Severity Discharge (ISD) criteria set. The ISD, developed by Interqual, is the grandparent of the AEP and the SMI. An historical review of the development of the AEP indicates that it was based on the ISD criteria and subsequently modified by the AEP instrument developers. Similarly, the SMI criteria set was based on and adapted from both the ISD and the AEP criteria. Questions concerning the relative reliability and validity of the ISD criteria remain unanswered. The ISD appears to be the utilization review criteria most widely used by Michigan hospitals, utilization review committees and organizations, and by quality assurance and assessment organizations (informal observation). One of the reasons that independent assessment of the ISD has not occurred is, in part, due to the lack of a paper and pencil score sheet designed to aid reviewers in the determination of the appropriateness of hospitalization. Because of the importance of such UR criteria, and because of the wide use of the ISD criteria set, Strumwasser, Paranjpe, Share, Sell and Stump (in progress) are conducting a study, funded by the Michigan Health Care Education and Research Foundation, to examine the reliability and validity of the ISD (final report due for distribution November, 1987). Preliminary results indicate that the ISD may, in some instances, be both more reliable and valid than the AEP (especially the ISD elective surgery criteria which are more valid than the SAEP).

A final comment on the validity of these criteria is appropriate. While we have covered the appropriate and inappropriate use and recommended cautions, we also indicate when appropriately used, these criteria yield valuable information, especially in use as assessment and screening tools. **These criteria should never be used when decisions regarding hospital care are being considered without the active involvement of well trained and highly qualified physician consultants.** These criteria sets, whether used manually or used in the context of a computerized program, have definite limits to their validity. Under no circumstances should they be considered stand alone decision instruments. The data on validity indicate quite convincingly that the upper limit of the validity of the AEP appears to be approximately 70%. That is, even under the best of circumstances, these criteria are confirmed by physician judgment in only 7 out of 10 instances of non-acute finding. This finding holds under the most liberal set of circumstances and across all organizational settings (i.e., FFS and HMO physicians). The AEP is valid at most 70% whether the validity confirmation is conducted by fee-for-service or HMO physicians. This level of validation is based on instances in which at least one physician, from either setting, was needed to corroborate an instrument finding of non-acute care. It is difficult to argue with this upper limit of validity since within our methodology we employed corroboration by HMO physicians employed by Kaiser Permanente Medical Group, of Kaiser Permanente Medical Center in Oakland, California. This particular HMO was chosen because in published studies, their rate of hospitalization is the lowest per capita of any HMO in the nation, and because the west coast, and California in particular, has been found to have one of the lowest rates of hospitalization, per capita, of any state in the nation. Thus, if any physician considered

care to be non-acute, it is more than likely that we would be able to find at least one of the three physicians from the HMO setting indicating care to be non-acute. When we match HMO physician assessments with the criteria assessments, we find that even the HMO physicians requiring at least one out of three corroboration yields a validity coefficient of approximately 80%. This finding raises an important question: Can the validity of these criteria, under lenient assumptions, be improved beyond 80% or is the nature of the decision regarding hospitalization so difficult and so diverse that validation or agreement by physicians with the criteria is unlikely to surpass the 80% level? It is our opinion that the criteria can be improved. In a very general sense, one way of improving the validity of the criteria is to find less care rather than more care non-acute. Our findings have convincingly demonstrated that, when RNs using the criteria make an error in non-acute assessment, the error almost always tends to be a false-positive identification of non-acute care. The developers of the AEP criteria need to refine their instruments in an effort to determine the kinds of cases, the circumstances, and the situations under which the AEP criteria overestimate the extent of non-acute care. It is likely that the AEP is more valid on certain kinds of admissions than others. Similarly, there may be a variety of circumstances in which these instruments are especially poor predictors or assessors of non-acute care. In the meantime, special caution toward over-estimation of non-acute care is critical. [It should be noted that in refining the AEP to reduce the false-positive identification of non-acute care, there are trade-offs between false-positives and false-negatives and that the two errors are probably not equally problematic. If there is an inexpensive review mechanism with high numbers of false-positive findings, perhaps (depending on cost/benefit analysis), a large

number of them can be tolerated to avoid a large number of false-negatives (as long as 100% of the "false-positives" are reviewed by physicians) (Beebe, personal communication, undated, parantheses added).]

Finally, it is our opinion that in all written material and presentations by developers and users of these criteria, an appropriate caution or warning be strongly and prominently displayed. The user should be warned that these criteria have been found to overestimate the extent of non-acute hospitalization. **Overestimation may negatively affect the quality of health care and the assessment of hospital, physician and community specific rates of non-acute care. Independent physician confirmation of non-acute AEP findings is required before non-acute evaluation is considered final.**

CHAPTER V

REFERENCES

REFERENCES

Blumenfeld, W.S. Critique of (1) "Methods to Determine Inappropriate Use of Hospital Services: The Appropriateness Evaluation Protocol (AEP)" and (2) "AEP Instructor's Manual Technical Note #5 (Construct Validity) and Technical Note #6 (Content Validity)". Report submitted to Sherry A. Terrell, Health Care Financing Administration, April, 1983.

Blumenfeld, W.S. Critique of the development and evaluation of the Standardized Medreview Instrument. Report submitted to Sherry A. Terrell, Health Care Financing Administration, March, 1985.

Browning, F.E. The record in hospital bed utilization review: A handbook for the medical staff, 1965, Chicago: American Medical Association, Dept. of Hospital and Medical Facilities.

Browning, F.E., and Crump, L. Report to the Patient Care Planning Council on a bed utilization study. Rochester, N.Y. Council on Social Agencies, 1969 (mimeograph).

Cohen, J.A. A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 1960, 20 (1), 37-42.

Gertman, P.M., and Bucher, B.M. Inappropriate hospital bed days and their relationship to length of stay parameters. Presented at the Medical Care Section, 99th meeting of the American Public Health Association, Minneapolis, Minnesota, October, 1969.

Gertman, P.M. and Restuccia, J.D. The Appropriateness Evaluation Protocol: A technique for assessing unnecessary days of hospital care. Medical Care, 1981, 19 (8), 855-871.

Gibson, R., Waldo, D.R. and Levit, K.R. National Health Expenditures, 1982. Health Care Financing Review, 1983, 5 (1), Fall, 1-31.

Hess, I., Riedel, D.C. and Fitzpatrick, T.B. Probability sampling of hospitals and patients. Ann Arbor: Health Administration Press (Second Edition), 1975.

Luft, H.S. Health Maintenance Organizations: Dimensions of Performance. New York: John Wiley and Sons, 1981.

Moorehead, M.A., Donaldson, R., et al. A study of the quality of hospital care secured by a sample of Teamsters' family members in New York City. N.Y.: Columbia University, School of Public Health and Administrative Medicine, 1964.

Querido, A. Analysis of length of stay. The efficiency of medical care. Leiden: H.E. Stenfert Kroese, NV., 1963.

Restuccia, J.D., Gertman, P.M., Dayno, S.J., Kreger, B.E. and Lenhart, G.M. The appropriateness of hospital use. Health Affairs, 1984 (Summer), 130-138.

Restuccia, J.D. and Holloway, D.C. Barriers to appropriate utilization of an acute care facility. Medical Care, 1976, 14 (7), 559-573.

Rishpon, S. Lubacsh, S. and Epstein, L.M. Reliability of a method of determining the necessity for hospitalization days in Israel. Medical Care, 1986, 24(3), 279-282.

Strumwasser, I., Paranjpe, N.V. and Hall, H. Determining Non-acute Hospital Stays. Business and Health, 1987, 4 (4), February, 18-22.

Strumwasser, I. and Paranjpe, N.V. Estimates of Non-Acute Hospital Care: "Then (1983) and Now (1986)", in progress.

Strumwasser, I., Paranjpe, N.V., Share, D., Sell, L. and Stump, M. "A Comparative Analysis of the Reliability and Validity of the AEP, SMI and ISD", in progress.

SysteMetrics (Moynihan, C., et al.) The valid and reliable measurement of non-acute hospital utilization in a nationally representative sample. Final report, HCFA Contract No. 500-80-0053, February 25, 1984.

Zimmer, J.G. Length of stay and hospital bed misutilization. Medical Care, 1974, 12 (5), 453-462.

Zimmer, J.G. An evaluation of observer variability in a hospital bed utilization study. Medical Care, 1967, 5 (4), 221-233.

Zimmer, J.G., and Groomes, E.W. An observer reliability study of physicians (sic) and nurses (sic) decisions in utilization review of chronic care facilities. Medical Care, 1969, 7 (1) 14-20.

CHAPTER VI

ATTACHMENTS

ATTACHMENT A

APPROPRIATENESS EVALUATION PROTOCOL

IENT'S NAME _____ FORM CONTROL # _____
(LAST) (FIRST) (INITIAL)

SPITAL NAME _____ PATIENT'S RECORD # _____

APPROPRIATENESS EVALUATION PROTOCOL

CARD 1
1/1

1/1

2-6/

m Control # _____ Year ____ ____ 7-8/
(2-6) 7 8

pital Name _____ Hospital Code # _____ 9-11/
(9-11)

ent's Age (Last Birthday) ____ ____ (14)
12 13

Patient's Sex (check) M ( ) 1 12-13/
F ( ) 2

14/

nary Insurance Coverage:
eck one only)

(15)

dicare ( ) 1
dicaid ( ) 2
e Cross ( ) 3
mmercial ( ) 4
er Third Party ( ) 5
-Pay ( ) 6

Hospital Service: (check) on admission:

(16)

Medicine ( ) 1
Surgery ( ) 2
Gynecology ( ) 3
Pediatrics ( ) 4
Psychiatry ( ) 5
Obstetrics ( ) 6
Other _____ ( ) 7

on date being reviewed:

(17)

Medicine ( ) 1
Surgery ( ) 2
Gynecology ( ) 3

15/

16/

17/

mission Date ☐☐/☐☐/☐☐ 18-22/
(18-22) Month Day Year

Date being reviewed ☐☐/☐☐/☐☐ 23-27/
(23-27) Month Day Year

ngth of Stay (days) ___ 28-30/
(28-30)

Diagnoses, active, this admission:

a. _____ ICD-9-CM ☐☐☐.☐☐ 31-35/

b. _____ ICD-9-CM ☐☐☐.☐☐ 36-40/

c. _____ ICD-9-CM ☐☐☐.☐☐ 41-45/

Major Procedures, this admission, on or before day reviewed?

a. Therapeutic

1. _____ ICD-9-CM ☐☐.☐☐ 46-49/

2. _____ ICD-9-CM ☐☐.☐☐ 50-53/

b. Diagnostic

1. _____ ICD-9-CM ☐☐.☐☐ 54-57/

2. _____ ICD-9-CM ☐☐.☐☐ 58-61/

Reason for Admission (Optional) : _____

_____ 62-63/

CARD 1
CONT.

riteria of Admission Appropriateness

Severity of Illness Criteria

		NO	YES	
1.	Sudden onset of unconsciousness or disorientation (coma or unresponsiveness).	(0) ___	(1) ___	64/
2.	Pulse Rate : A. Less than 50 per minute. B. Greater than 140 per minute.	(0) ___	(1) ___	65/
3.	Blood Pressure : A. Systolic less than 90 or greater than 200 mm. Hg. B. Diastolic less than 60 or greater than 120 mm. Hg.	(0) ___	(1) ___	66/
4.	Acute loss of sight or hearing.	(0) ___	(1) ___	67/
5.	Acute loss of ability to move body part.	(0) ___	(1) ___	68/
6.	Persistent fever equal to or greater than 100 (p.o.) or greater than 101(R) for more than 5 days.	(0) ___	(1) ___	69/
7.	Active bleeding.	(0) ___	(1) ___	70/
8.	Severe electrolyte/Blood gas abnormality (any of the following) : A. Na < 123 mEq/L Na > 156 mEq/L B. K < 2.5 mEq/L K > 6.0 mEq/L C. CO_2 combining power (unless chronically abnormal) < 20 mEq/L CO_2 combining power (unless chronically abnormal) > 36 mEq/L D. Blood ph < 7.30 Blood ph > 7.45	(0) ___	(1) ___	71/
9.	Acute or progressive sensory, motor, circulatory or respiratory embarrassment sufficient to incapacitate the patient (inability to move, feed, breathe, etc.). Note: Must also meet Intensity of Service criterion simultaneously in order to certify. Do not use for back pain.	(0) ___	(1) ___	72/
10.	EKG evidence of acute ischemia; must be suspicion of a new MI.	(0) ___	(1) ___	73/
11.	Wound dehiscence or evisceration.	(0) ___	(1) ___	74/

CARD 2
1/2

1/2
2-6/

B. Intensity of Service

		NO	YES	
1.	Intravenous medications and/or fluid replacement (does not include tube feedings).	(0) ___	(1) ___	7/
2.	Surgery or procedure scheduled within 24 hours requiring: A. General or regional anesthesia or B. Use of equipment, facilities, procedure available only in a hospital.	(0) ___	(1) ___	8/
3.	Vital sign monitoring every 2 hours or more often (may include telemetry or bedside cardiac monitor).	(0) ___	(1) ___	9/
4.	Chemotherapeutic agents that require continuous observation for life threatening toxic reaction.	(0) ___	(1) ___	10/
5.	Treatment in an I. C. U.	(0) ___	(1) ___	11/
6.	Intramuscular antibiotics at least every 8 hours.	(0) ___	(1) ___	12/
7.	Intermittent or continuous respirator use at least every 8 hours.	(0) ___	(1) ___	13/

OVERRIDE OPTIONS

8. Other services justifying appropriateness :
   Yes 1( ) Description: _____ 14/
   (14) (15–16) 15–16/
   _____

9. Criteria met, but inappropriate nevertheless :
   Yes 1( ) Description: _____ 17/
   (17) (18–19) 18–19/
   _____

	NO	YES	
Appropriateness of admission:	(0) ___	(1) ___	20/
Decision made solely on basis of override:	(0) ___	(1) ___	21/

Reason for admission failing to meet Criteria of Appropriateness (see list of reasons):

Reason 1 _____ 22–23/
Reason 2 _____ 24–25/
Reason 3 _____ 26–27/

Criteria of Appropriateness of Day of Care

CARD 2
CONT.

A. Medical Services

	NO	YES
1. Procedure in operating room that day.	(0) ___	(1) ___ 28/
2. Scheduled for procedure in operating room the next day, requiring pre-operative consultation or evaluation.	(0) ___	(1) ___ 29/
3. Cardiac catheterization that day.	(0) ___	(1) ___ 30/
4. Angiography that day.	(0) ___	(1) ___ 31/
5. Biopsy of internal organ that day.	(0) ___	(1) ___ 32/
6. Thoracentesis or paracentesis that day.	(0) ___	(1) ___ 33/
7. Invasive CNS diagnostic procedure (e.g., lumbar puncture, cysternal tap, ventricular tap, pneumoencephalography) that day.	(0) ___	(1) ___ 34/
8. Any test requiring strict dietary control, for the duration of the diet.	(0) ___	(1) ___ 35/
9. New or experimental treatment requiring frequent dose adjustments under direct medical supervision.	(0) ___	(1) ___ 36/
10. Close medical monitoring by a doctor at least three times daily (observations must be documented in record).	(0) ___	(1) ___ 37/
11. Post-operative day for any procedure covered in numbers 1, or 3-7 above.	(0) ___	(1) ___ 38/

OVERRIDE OPTIONS

12. Other services justifying appropriateness :
Yes 1( ) Description: _____ 39/
(39) (40–41)
_____ 40–41/

13. Criteria met, but inappropriate nevertheless :
Yes 1( ) Description: _____ 42/
(42) (43–44)
_____ 43–44/

B. Nursing/Life Support services:

	NO	YES
1. Respiratory care – intermittent or continuous respirator use and/or inhalation therapy (with chest PT, IPPB) at least thrice daily.	(0) ___	(1) ___ 45/
2. Parenteral therapy – intermittent or continuous IV fluid with any supplementation (electrolytes, protein, medications).	(0) ___	(1) ___ 46/
3. Continuous vital sign monitoring, at least every 30 minutes, for at least four hours.	(0) ___	(1) ___ 47/
4. IM and/or SC injections at least twice daily.	(0) ___	(1) ___ 48/
5. Intake and output measurement.	(0) ___	(1) ___ 49/
6. Major surgical wound and drainage care (chest tubes, T-tubes, hemovacs, Penrose drains).	(0) ___	(1) ___ 50/
7. Close medical monitoring by nurse at least 3 times daily, under doctor's orders.	(0) ___	(1) ___ 51/

CARD 2
CONT.

OVERRIDE OPTIONS

8. Other services justifying appropriateness :
   Yes 1( ) Description: _____ 52/
   (52) (53--54)
   _____ 53--54/
9. Criteria met, but inappropriate nevertheless :
   Yes 1( ) Description: _____ 55/
   (55) (56--57)
   _____ 56-57/

C. Patient Condition

	NO	YES	
Within 24 hours on or before day of review:			
1. Inability to void or move bowels (past 24 hours) not attributable to neurologic disorder.	(0) ___	(1) ___	58/
Within 48 hours on or before day of review:			
2. Transfusion due to blood loss.	(0) ___	(1) ___	59/
3. Ventricular fibrillation or ECG evidence of acute ischemia, as stated in progress note or in ECG report.	(0) ___	(1) ___	60/
4. Fever at least 101 rectally (at least 100 orally), if patient was admitted for reason other than fever.	(0) ___	(1) ___	61/
5. Coma – unresponsiveness for at least one hour.	(0) ___	(1) ___	62/
6. Acute confusional state, not due to alcohol withdrawal.	(0) ___	(1) ___	63/
7. Acute hematologic disorders, significant neutropenia, anemia, thrombocytopenia, leukocytosis, erythrocytosis, or thrombocytosis, yielding signs or symptoms.	(0) ___	(1) ___	64/
8. Progressive acute neurologic difficulties.	(0) ___	(1) ___	65/
Within 14 days before day of review:			
9. Occurrence of a *documented, new* acute myocardial infarction or cerebrovascular accident (stroke).	(0) ___	(1) ___	66/

OVERRIDE OPTIONS:

10. Other services justifying appropriateness :
    Yes 1( ) Description: _____ 67/
    (67) (68–69)
    _____ 68–69/
11. Criteria met, but inappropriate nevertheless :
    Yes 1( ) Description: _____ 70/
    (70) (71–72)
    _____ 71–72/

CARD 2
CONT.

VIII. **Appropriateness of Day Reviewed (check one) :**

_____ (1) Criteria met - day appropriate 73/

_____ (2) Criteria met but based on an override the day is deemed inappropriate

_____ (3) No criteria met - day inappropriate

_____ (4) No criteria met but based on an override the day is deemed appropriate

If the answer to the above is 2 or 3 , complete the following:

IX. **If day is inappropriate; is continued hospitalization necessary on medical grounds?** (0) No _____ (1) Yes _____ 74/

CARD 3
1/3

1/3
2-6/

X. **Reasons for a day failing to meet Criteria of Appropriateness:**
(see list of reasons)

A. For patients who need to be hospitalized:

Reason 1 _____ (7– 8) 7– 8/

Reason 2 _____ (9–10) 9–10/

Reason 3 _____ (11–12) 11–12/

13–14/

B. For patients who do not need further hospitalization: 15–16/

Reason 1 _____ (13–14) Responsibility _____ (15–16) 17–18/

Reason 2 _____ (17–18) Responsibility _____ (19–20) 19–20/

Reason 3 _____ (21–22) Responsibility _____ (23–24) 21–22/

23–24/

C. If "other" categories are used specify reason and/or response here: _____ 25–26/

_____ 27–28/

(25–26) (27–28)

Administrative Information

Field Reviewer Code _____ 29–31/
(29–31)

Central Office Reviewer Date _____

REASONS LIST FOR RETROSPECTIVE REVIEW

Reasons for Inappropriate Admission

01. Patient needs no institutional care: any needed diagnosis and treatment can be handled on an outpatient basis
02. Patient needs institutional care at a level other than an acute care hospital -- general (unspecified)
03. Patient needs care in a chronic disease hospital
04. Patient needs care in a skilled nursing home
05. Patient needs care in a non-skilled nursing home
06. Premature admission (e.g., on Friday for a procedure booked for the following Tuesday)
09. Other (specify)

Reasons for Inappropriate Acute Day of Stay

A. For patients who need continuing acute hospitalization:

10. Patient awaiting procedure in operating room
20. Patient awaiting performance of a diagnostic test or non-OR procedure
30. Awaiting result of a diagnostic test or consultation
49. Other (specify)

B. For patients who do not need continuing acute hospitalization:

50. Patient needs no further institutional care at any level

60. Patient needs lower level institutional care

69. Other (specify)

Responsibility Categories

70. Physician or hospital responsibility -- general (unspecified)

71. Inadequate discharge planning by physician or hospital
72. Overconservative medical management
79. Other (specify)

80. Patient or family responsibility, e.g., patient or family insists on patient's remaining in hospital

90. Environmental responsibility -- general (unspecified)

91. Patient from unhealthy environment is kept until environment becomes acceptable or alternative facility is found
92. Patient is convalescing from an illness, and it is anticipated that his/her stay in an alternate facility would be less than 72 hours
93. Bed unavailable at alternate facility
94. Transportation unavailable to alternate facility
99. Other (specify)

ATTACHMENT B

STANDARDIZED MEDREVIEW INSTRUMENT (SMI)

PATIENT BACKGROUND INFORMATION

uential # (2-5)

Rater # (6-7)

Hospital # (8-10)

Medical Record # (11-22)

CARD 1 BEG 1/1

ADMIT DATE (28)
99 = Missing

ADMIT TIME (30)
Hours in Military. Drop minutes.
99 = Missing

DISCHARGE DATE (36)
99 = Missing

DISCHARGE TIME (38)
Hours in Military. Drop minutes.
99 = Missing

DATE OF BIRTH (-44)
99 = Missing
Check if birthyear is an approximation ✓ = 1

SEX
1 = Male
2 = Female
3 = Not Available

EXPECTED SOURCE OF PAYMENT PRINCIPAL (7-48)
1 = Medicare
2 = Medicaid
3 = Other Govt.
4 = Workers Compensation
5 = Blue Cross
6 = Commercial Insurance
7 = HMO
8 = Self Pay Only
9 = No Charge
10 = Other
11 = Not Available

PRIMARY PLACEMENT (9-50)
1 = Medical Services—includes medical subspecialties. (See manual for list.)
2 = Surgical Service—includes surgical subspecialties. (See manual for list.)
3 = Psychiatric Unit
4 = Intensive Care/Coronary Care Unit
5 = Burn Care Specialty Unit
6 = Alcoholism/Chemical Dependency Unit
7 = Organized Rehabilitation Unit
8 = Trauma/Shock Unit
9 = Other
10 = Not Available

9. SOURCE OF ADMIT (51)
1 = Emergency Room
2 = Transferred/Admitted from Ambulatory Surgery Unit
3 = Transferred from other acute facility
4 = Transferred from skilled nursing facility
5 = Routine—(Physician's office, clinic, etc.)
6 = Not Available

10. DISPOSITION OF PATIENT (52)
1 = Routine/Home
2 = Transferred/referred to other acute facility
3 = Transferred/referred to skilled nursing home
4 = Transferred/referred to other non-acute facility. (Intermediate Care, Psychiatric Rehabilitation, Board and Care.)
5 = Referred to organized home care program
6 = Left AMA (against medical advice)
7 = Died
8 = Not Available

11. DIAGNOSES
ICD-9-CM CODE
Principal (53-57)
Other (58-62)
(63-37)

12. MAJOR PROCEDURES (Class I operations)
ICD-9-CM CODE Date Performed
Principal
Card 2 (8-11/12-17)
Other (18-21/22-27)
(28-31/32-37)
(38-41/42-47)

2-5/____ ____ ____ ____
6-7/____ ____
8-10/____ ____ ____
11-22/____ ____ ____ ____
____ ____ ____ ____
____ ____ ____ ____
23-28/____ ____ ____ ____ ____ ____
29-30/____ ____
31-36/____ ____ ____ ____ ____ ____
37-38/____ ____
39-44/____ ____ ____ ____ ____ ____
45/____
46/____
47-48/____ ____
49-50/____ ____
51/____
52/____
53-57/____ ____ ____ ____ ____
58-62/____ ____ ____ ____ ____
63-67/____ ____ ____ ____ ____

CARD 2 BEG 1/2

2-5/____ ____ ____ ____
6-7/____ ____
8-11/____ ____ ____ ____
12-17/____ ____ ____ ____ ____ ____
18-21/____ ____ ____ ____
22-27/____ ____ ____ ____ ____ ____
28-31/____ ____ ____ ____
32-37/____ ____ ____ ____ ____ ____
38-41/____ ____ ____ ____
42-47/____ ____ ____ ____ ____ ____

CARD 2 END

TURN PAGE FOR ADMISSION REVIEW

ADMISSION REVIEW

CRITERIA FOR ADMISSION: Select the primary criterion met, then go to Section II

(8-10)

LABORATORY ABNORMALITIES:

BLOOD/SERUM:

1. Acid Phosphatase increased (guidelines in manual)
2. ADH decreasing, with Polyuria (guidelines in manual)
3. Amylase >50% above lab normal (guidelines in manual)
4. Bilirubin >3 mg% (total serum bilirubin)
5. Blast Cells in peripheral blood
6. Blood Sugar: <40 or >400; or, fasting: >250
7. BUN >40 OR Creatinine >2
8. Calcium <7 or >13
9. Cortisol >3 times lab normal (guidelines in manual)
10. CO_2 combining power (HCO_3) <16 or >36
11. Hemoglobin (HGB) <6.0 or >18.0
12. Hematocrit (HCT) <25 or >55
13. Lithium Carbonate level above maximum normal (>1.5 units/ml)
14. pH <7.30 or >7.45
15. Platelet count <20,000 or >1,000,000
16. Potassium (K) <2.5 or >6.0
17. pO_2 <60
18. pCO_2 >50

__ 19. Prothrombin Time >20 seconds or 2 times >lab normal (not on anticoagulants)

__ 20. Sodium (Na) <123 or >150

__ 21. T_4 <2 or >20

__ 22. WBC <2,000 or >15,000

BACTERIOLOGY:

__ 23. Bacteria
- By Culture: Any site
- By Gram Stain: Spinal Fluid/Sputum
- By Colony Count: Urine remains >100,000 after 3 days antibiotic treatment

__ 24. Fungus
- By Culture/Gram Stain: Spinal fluid

OTHER LABORATORY FINDINGS:

__ 25. Spinal Fluid: RBC >5 or WBC >5 or sugar 35% below blood sugar level or protein above 45% or presence of xanthochromia

__ 26. Toxic level of drug or other agent (guidelines in manual)

__ 27. Urinary Amylase, 2 hour, elevated (guidelines in manual)

__ 28. VMA >9 (24 hour urine)

MEDICAL PROBLEMS/DISEASES/COMPLICATIONS OF MEDICAL OR SURGICAL CARE:

__ 29. Abscess: Lung, Gastrointestinal, Brain, Liver, Peritonsillar, or Multiple Skin

__ 30. Abuse: Alcohol or drug

__ 31. Arthritis: Polyarthritis, acute (i.e., acute rheumatoid arthritis or systemic lupus or acute rheumatic fever); or, Septic Arthritis (acute invasive or infectious process of bone or joint)

__ 32. Allergic Reaction, (severe)

__ 33. Baker's cyst, ruptured

__ 34. Burns

__ 35. Breakdown of Colostomy site

__ 36. † Calculus/Stone in urinary tract

__ 37. † Cardiac Pathology

__ 38. Cellulitis, widespread; or, localized with vascular insufficiency

__ 39. † Coma

__ 40. Complications of Dialysis/Fistula/Shunt

__ 41. Compression Entrapment Neuropathy: (Excludes carpal or tarsal tunnel syndromes. If admitted for carpal or tarsal tunnel release, indicate by using Section II, item 2.)

__ 42. Conjunctivitis, gonorrheal

__ 43. † CVA/Embolic or Thrombotic Phenomenon

__ 44. † Diabetes, WITH pregnancy (for medication adjustment) or WITH uncontrolled infection

__ 45. Endophthalmitis

__ 46. Extravasation of radiological contrast material

__ 47. Fistula, A-V (first admission for or for surgery)

__ 48. Foreign Body or Aspiration

__ 49. Fractures (includes pathological)

__ 50. Glaucoma: Acute angle closure or pressure >30

__ 51. Head Trauma, recent

__ 52. Herniated Intervertebral Disc

__ 53. Infarction: Bowel, Brain, Cerebrovascular, Myocardial, Pulmonary

__ 54. Infection: Periorbital/Intraocular

__ 55. Inflammation, Pelvic (PID), persistent after 3 days antibiotic treatment

__ 56. † Labyrinthitis

__ 57. Lens Dislocation

__ 58. † Lung: Atelectasis (collapse), abscess, hemothorax, pneumothorax or air in mediastinum

__ 59. Malfunctioning gastrostomy tube

__ 60. Malfunctioning pacemaker, suspected or actual

__ 61. † Meniere's Disease

__ 62. † Multiple Sclerosis (new diagnosis or acute exacerbation)

__ 63. Organ Transplant Patient with possible or probable rejection

__ 64. Osteomyelitis

__ 65. † Otitis Media/Mastoiditis with possible intracranial complications

__ 66. Pneumonia, bacterial (excluding pneumococcal) or as modified by * list

__ 67. Pregnancy: ectopic (including suspected); complicated; or, for normal delivery

__ 68. Psychopathology

__ 69. † Psoriasis, generalized

__ 70. Retinal Detachment

__ 71. Spinal Cord Injury, acute or suspected

__ 72. Status Asthmaticus

__ 73. Trauma, other

__ 74. Urinary tract shunt breakdown or change

__ 75. Wound: disruption, dehiscence, or post-operative infection

__ 76. Wound, open: traumatic

*See explanatory notes † acute or for management

ADMISSION REVIEW CONTINUED ON NEXT PAGE

CRITERIA FOR ADMISSION (continued)

ABNORMAL SIGNS:

77. Abdomen: Rigidity or rebound tenderness WITH: Temp. >100°F or nausea and vomiting or WBC >10,000
78. † Ascites
79. Babinski, positive (first occurrence)
80. Bowel distention evidenced by air fluid levels on radiology
81. † Claudication, intermittent (for definitive treatment, management or evaluation)
82. Cyanosis
83. Carotid Bruit WITH TIA or Aphasia or Ataxia or Syncope
84. Colic, Renal, severe with suspicion of stone or obstruction
85. Enlargement or block, progressive, of ventricle (brain)
86. † Epidermolysis/primary skin disease with defoliation
87. Gangrene
88. Hemarthrosis
89. Hydronephrosis/Hydroureter
90. Increased Intracranial or Spinal Fluid Pressure
91. † Mass/Tumor/Lesion
92. Obstruction/Occlusion
93. Output Urine Below 20 cc./hr or 400 cc./24 hrs.
94. Pap Smear, Class IV
95. Papilledema
96. Perforation of GI tract with radiological evidence
97. † Pleural Effusion/Pulmonary congestion, edema or fluid
98. Retention of urine over 12 hours requiring catheterization
99. Sella Turcica enlargement
100. Sensory or motor deficit with abnormal neurological reflex
101. Shock (BP <100/40)
102. Spinal fluid discharge from ear or nose/ Cerebrospinal Otorrhea
103. Splenomegaly/Spleen enlargement (first occurrence)
104. Tuberculosis (includes suspected) with fever, weight loss AND compatible chest x-ray
105. Thrombophlebitis
106. Ulcer: Draining, not responsive to outpatient management, for treatment; or severe corneal or herpetic ulcer
107. Unconsciousness
108. Urethral Stricture or Dilation
109. Urinary Extravasation
110. Vital Sign Abnormality

SYMPTOMS:

111. Bleeding/Hemorrhage, gross or uncontrolled
112. † Dermatitis, generalized, severe
113. Eruption of skin: diffuse vesicular or bullous
114. † Functional impairment, physical
115. Pain (severe, incapacitating)
116. † Seizure Activity
117. Urine leakage into vagina/rectum/colon

ADMISSION EVALUATION (check one):

PPEARS JUSTIFIED BY CRITERIA AND/OR RECORD REVIEW

(1-14)

☐ = 1000 Criterion met

☐ = 2000 No criterion met, but admission was for operation/procedure which requires acute hospital setting (see Manual Appendix A)

☐ = 3____ No criterion met, but need for admission supported by logical extension of criterion #____. (Explain below)

☐ = 4000 No criterion met, but admission appears justified on the basis of record review. (Explain below)

PPEARS NOT JUSTIFIED. (Requires Reason Code from list 1)

REASON CODE (15-16) (17-18) (19-20)

☐ = 5000 No criterion met.

☐ = 6000 Criterion met, but record review does not support need. (Explain below)

XPLANATION _____

_____

_____

O NOT WRITE BELOW THIS LINE

(80)

CARD 3 BEG

1/3

2-5/____

6-7/____

8-10/____

11/14____

15-16/____

17-18/____

19-20/____

21-80/see box

CARD 4 END

TURN PAGE FOR REVIEW OF INDIVIDUAL DAYS OF CARE: LEVEL OF SERVICE

EXPLANATORY NOTES FOR ADMISSION CRITERIA

This symbol, in front of a criterion statement, means the condition must be ACUTE, or an exacerbation of the chronic condition, or the admission must be for definitive treatment, management or evaluation of the condition.

MEDICAL PROBLEMS/DISEASES/COMPLICATION OF MEDICAL OR SURGICAL CARE

30. ABUSE, ALCOHOL OR DRUG
- Acute alcohol withdrawal symptoms
- Alcoholic coma or severe stupor
- Delirium Tremens
- Drug Detoxification
- †Hallucinations (visual or auditory)
- Seizures (withdrawal or toxic)
- †Uncontrolled or excessive drug abuse

32. ALLERGIC REACTION (Severe)
- Anaphylactic Shock
- Angioneurotic Edema
- †Giant Urticaria (for definitive treatment, management or evaluation)

34. BURNS
- Burns of the Eye
- 3° Burns
- 2° Burns of >9% of the body, or of the face
- Chemical burns of mouth or throat
- Pulmonary Burns, including suspected
- Any burn requiring graft or ICU/Burn Unit Care

37. † CARDIAC PATHOLOGY (Acute Onset)
- Myocardial Infarction or Ischemia
- Arrhythmias: Fibrillation, Flutter, Tachycardia (>120), Bradycardia (<50)
- 2° or 3° Heart Block
- Pericardial Friction Rub
- Digitalis Toxicity
- Congestive Heart Failure
- Massive Cardiac Enlargement (must be defined by physician or radiologist as massive and will most likely be evidenced by x-ray)

39. † COMA
- Diabetic Coma
- Alcoholic Coma
- Hepatic Coma (or liver flap with altered level of consciousness)

40. COMPLICATIONS OF DIALYSIS/FISTULA/SHUNT
- Clot in shunt
- Occluded shunt
- Swelling at fistula site
- Fistula Leakage
- Fistula Obstruction

43. † CVA/EMBOLIC OR THROMBOTIC PHENOMENON
- Includes:
  - Basilar artery syndrome
  - Carotid artery syndrome
  - Transient Ischemic Attacks (TIA)
  - Vertebral artery syndrome
- Excludes: Long-standing residual neurological deficits of a CVA

48. FOREIGN BODY OR ASPIRATION (Excludes: long-standing not admitted for definitive treatment, management, or evaluation)
- Acute injury with presence of foreign body, (e.g., bullet or other projectile)
- Aspiration of foreign material
- Foreign body in bladder/urethra/bronchial tree/ intraocular or intraorbital space
- Ingestion or inhalation of caustic substance (sodium hydroxide, Drano, Liquid Plumber, sodium carbonate [found in dishwashing detergents], ammonia, lye, household bleaches [Clorox], potassium hydroxide, acids [nitric, hydrochloric, sulfuric, carbolic])
- Smoke Inhalation

49. FRACTURES (Includes Pathological)
- Fracture with possibility of vascular, sensory or motor compromise
- Open or Compound Fracture
- Radiological evidence of fracture of: Spine, Femur, Pelvis, 3 or more ribs, sternum, skull, jaw, facial bone, pubic rami, tibia, ankle, multiple extremities (any weight bearing area of body)
- Colles' Fracture in elderly (>65) persons

★ 51. HEAD TRAUMA, RECENT
- Recent head trauma WITH vomiting OR increasing blood pressure OR fluctuating level of consciousness OR pulse less than 60
- For 24 hour observation

★ 52. HERNIATED INTERVERTEBRAL DISC
- Disc Protrusion
- Filling Defect on myelogram
- With acute neurologic deficit

★ 66. PNEUMONIA
- Any Bacterial Pneumonia, except Pneumococcal
- Pneumococcal or any other type of pneumonia (e.g., viral, allergic, etc.) in the presence of one of the following conditions:
  - Age — very young (<5) or very old (>65)
  - General debilitated condition of patient
  - Presence of Diabetes Mellitus
  - Outpatient treatment failure

★ 67. PREGNANCY: Ectopic (including suspected); for normal delivery; or complicated by:
- Abdominal rigidity or tenderness
- Bleeding, vaginal in 2nd or 3rd trimester
- Cephalopelvic disproportion
- Cervical dilation during gestation
- Diabetes, newly discovered or requiring medication adjustment
- Fetal distress
- Hyperemesis
- Hyperreflexia
- Incompetent cervix
- Premature labor
- Excessive sudden weight gain and edema
- Extrauterine palpable mass
- Leak of Amniotic fluid
- Placenta Previa (by ultrasound)
- Premature rupture of membranes
- Previous Cesarean Section (for delivery)
- Protrusion of fetal part from vagina
- Retarded fetal growth
- Pre-eclampsia with:
  1. Systolic blood pressure recently elevated more than 30 above normal; or,
  2. More than 15 above normal AND protein in urine; or,
  3. Diastolic pressure >100

★ 68. PSYCHOPATHOLOGY (use for patients admitted for psychiatric care only; those criteria for physical conditions of non-psychiatric patients, such as unconsciousness, are listed elsewhere in the criteria)
- A. Incapacitating Emotional States
  Anxiety or agitation, severe/incapacitating; Depression, severe; Mania
- B. Behavior Aberrations
  Assaultive behavior; Bizarre or delusional behavior; Inability to maintain nutrition (or refusal to do so); Self mutilative behavior; Suicide attempt/ideation
- C. Functional Impairment Due to Thought Processes
  Delirium, Disorientation or memory impairment to a degree endangering welfare; Disorganized thought process; Stuporous
- D. Complete Psychophysiological Dysfunction
  Rigidity, total body; Unconsciousness

★ 73. TRAUMA, OTHER
- Acute trauma to neck or throat interfering with swallowing and/or respiration; Trauma to urinary tract system; Genital trauma requiring surgical repair; Severe crushing injury; Suspicion of ruptured organ; Trauma with neurovascular deficit; Loss or damage of skin >10% of body surface

★ 76. WOUND, OPEN: Traumatic
- Includes: Acute loss of limb; Acute loss or portion of external genitalia; Infected or with sepsis; Penetrating wound of abdomen, chest cavity, or urinary tract system; Perforation or laceration of eyeball (any significant wound of the eye)
- Excludes: Open wounds of skin or multiple contusions or abrasions (If physician indicates admission is necessary for evaluation of another condition, as, "possible concussion," or "possible internal injuries," check the criterion that applies, as "Other Trauma" or "Head Trauma," etc.)

EXPLANATORY NOTES FOR ADMISSION CRITERIA (Continued)

ABNORMAL SIGNS:

★ **91. MASS/TUMOR/LESION:** Long-standing will be included here only if recurrence or change occurs, or if admitted for re-evaluation

Cancer — lung, osteogenic sarcoma, Ewing's tumors
Chest, newly discovered
Costovertebral angle mass (kidney area)
Intraocular orbital tumor
Thyroid mass, causing airway obstruction
Space occupying lesion of CNS
Abdominal Mass
Breast Mass
Testicular Mass
Urinary Tract
Mass in or on Tonsils

★ **92. OBSTRUCTION/OCCLUSION**

Larynx or pharynx, acute obstruction
Occlusion, central retinal artery
Peripheral Vascular Occlusion — evidenced by absence of pulses WITH distal ischemia (rest pain, loss of sensation, ulceration, gangrene)
Acute occlusion of vessel
Urethral obstruction
Gastrointestinal: nonpassage of contrast material

★ **100. SENSORY OR MOTOR DEFICIT WITH ABNORMAL NEUROLOGICAL REFLEX**

— of peripheral nerve origin (polyneuropathies; sensory loss, paralysis, loss of reflexes; Guillain-Barre syndrome)
— toxic neuropathies (uremia, alcohol)
— mononeuropathies (infections, as herpes zoster; or traumatic, as carpal tunnel syndrome or meralgia paresthetica)
— central nervous system lesions (CVA, stroke, tumor)

★ **105. THROMBOPHLEBITIS**

Deep thrombophlebitis
Calf swelling or x-ray documentation of thrombophlebitis

★ **110. VITAL SIGN ABNORMALITY**

Temperature: ≥100°F (37.8°C) with neutrophil count <2000
>102°F (38.9°C) with bacteria by culture/smear, or with WBC >12,000
Pulse: <50 or >140
Respiratory Rate: >32
Blood Pressure: Systolic <80 or >200
Diastolic >120

SYMPTOMS:

★ **111. BLEEDING, GROSS OR UNCONTROLLED**

Arterial bleeding
Eye: anterior chamber or vitreous hemorrhage
Hematoma: Intra-abdominal or Intracranial
† Hemoptysis, acute
Gastrointestinal bleeding, suspected or diagnosed, evidenced by:
Hematocrit <30
Hemoglobin <9.5
Hematemesis
Blood in gastric expirant
Gross rectal bleeding
Gross Hematuria, first admission for or with known urinary tract cancer or with calculus
Post-operative bleeding, uncontrollable

★ **114. †FUNCTIONAL IMPAIRMENT, PHYSICAL** — must be of recent onset, acute, or admitted for definitive treatment or evaluation

Sight or speech loss
Inability to move any body part or parts
Inability (physical) to maintain nutrition (cannot feed self)
Mental status causing physical impairment: (delirium, mania, disorientation, refusal to maintain nutrition)
Use only for non-psychiatric patients. For psychiatric patients, see "Psychopathology"

★ **115. PAIN** (severe, incapacitating)

Acute chest pain, dyspnea, or cyanosis
Pleuritic chest pain with friction rub
Recent onset of pelvic pain with pelvic mass OR temperature of >101°F (38.4°C)
Acute onset of severe testicular pain (for more than 3 hours prior to admission)
Painful sustained erection

★ **116. †SEIZURE ACTIVITY** — acute onset or exacerbation

Uncontrolled seizures
Initial/first seizure activity
Withdrawal or toxic (drug or alcohol or poison)

REVIEW OF INDIVIDUAL DAYS OF CARE: LEVEL OF SERVICE

CRITERIA FOR ACUTE-CARE SERVICES (36-49) Check the primary criterion met each day, then go to Section II	Date:	(8-11)	(12-15)	(16-19)	(20-23)	(24-27)	(28-31)	(32-35)
OPERATIONS/PROCEDURES		(36-37)	(38-39)	(40-41)	(42-43)	(44-45)	(46-47)	(48-49)
PROCEDURE PERFORMED: Class 1 only (see manual for list)	1.							
PRE-OP PREPARATION DAYS (Class 1 only): allow 1 day, except for bowel prep allow 2 days.	2.							
CONTINUOUS OR TIMED SERVICES:								
ARTERIAL BLOOD GASES DAILY	3.							
BLADDER IRRIGATION, CONTINUOUS	4.							
CATHETER: (excludes long-standing) Indwelling for acute urinary retention or hematuria; or, Ureteral catheter in place	5.							
DRAINS: Gastric/chest (underwater)/intestinal	6.							
ICU/CCU/BURN CARE UNIT	7.							
ISOLATION: Protective (Reverse or Psych); Respiratory	8.							
IV FLUIDS OR MEDS: Any, except to keep vein open (TKO)	9.							
MONITORING PROGRAM IN EFFECT	10.							
RESPIRATORY CARE	11.							
SUCTIONING, 4 TIMES DAILY: endotracheal, nasotracheal, transtracheal	12.							
TRACTION: skeletal, skin, pelvic, sternal, rib, or Crutchfield Tongs; plus 3 days	13.							
ACUTE SERVICES DEPENDENT ON CONDITION OF PATIENT								
CARDIOVERSION/DEFIBRILLATION	14.							
DIALYSIS, INITIAL COURSE: Peritoneal/Renal (Hemodialysis)	15.							
EYE CARE: Glaucoma (intraocular pressure) control (observation and medication), Retinal monitoring, Leaking ocular wound care, Topical meds hourly	16.							
HEMORRHAGE CONTROL: Arterial, nasal, post-op	17.							
HYPERBARIC OXYGENATION: Skin care, gas gangrene, severe infections	18.							
INTRAMUSCULAR (IM) INSULIN FOR NEWLY DIAGNOSED DIABETES OR WITH ACIDOSIS OR COMA	19.							
NASOGASTRIC TUBE OR GASTROSTOMY FEEDINGS	20.							
PSYCHIATRIC CARE (Includes Substance Abuse Care)	21.							
SKIN CARE: Surface Burn therapy or Ultraviolet daily	22.							
TRANSFUSION: Blood, whole or component/Bone Marrow	23.							
CLASS 2 PROCEDURE PERFORMED	24.							
OBSERVATION Treatment, management or evaluation regimen requiring close clinical observation and/or laboratory monitoring.								
LAB TESTING: Blood Sugar q4h, Insulin tolerance, Metapyrine, Regitine, Vasopressin	25.							
MEDICATION ADJUSTMENT OR REGULATION	26.							
GENERAL OBSERVATION	27.							
PATIENT CONDITION CHANGING OR UNSTABLE WHILE AWAITING SURGERY (Caution: patient must require stabilization prior to surgery: this must be well documented in the record. Do not use this criterion indiscriminately. Patient management regimen must be directed toward this stabilization.)	28.							
PATIENT EDUCATION AND TRAINING SERVICES								
APPLIANCE CARE: Salivary Cutaneous Fistula; Stoma care; Tracheostomy Care, Initial; Urinary Drain (If patient is not scheduled for closure or removal, allow 2 days maximum)	29.							
PHYSICAL THERAPY: Skilled PT or speech therapy twice daily; Begin rehabilitation evaluation, goal setting and management-program initiation; Participation in an organized rehabilitation program	30.							
DAILY EVALUATION LEVEL OF SERVICE APPEARS: (Check one below, then go to Section III)		(50-51)	(52-53)	(54-55)	(56-57)	(58-59)	(60-61)	(62-63)
ACUTE LEVEL: Criterion met	✓ = 31							
ACUTE LEVEL: No criterion met, but logical extension applies. (Enter criterion # and explain on attached page.)	✓ =							
ACUTE LEVEL: No criterion met, but record review supports. (Explain on attached page.)	✓ = 32							
NONACUTE LEVEL: No criterion met	✓ = 33							
NONACUTE LEVEL: Criterion met, but record review does not support. (Explain on attached page.)	✓ = 34							
Requires REASON CODE from List 2 → Most Likely (Required) (64-77)								
Second (Optional) card 5 (8-21)								
Third (Optional) (22-35)								
I. PHASE OF STAY Assign one phase number for each day of the stay. (Use categories below. See manual for clarification.)		(36)	(37)	(38)	(39)	(40)	(41)	(42)
Pretreatment Phases: 1. Workup in progress; 2. Treatment scheduled — Treatment Phases: 3. Surgical; 4. Medical; 5. Transitional — Post-treatment Phase: 6. Treatment completed								

CARD 4 BEG

1/4
2-5/____ ____ ____ ____
6-7/____ ____
8-11/____ ____ ____ ____
12-15/____ ____ ____ ____
16-19/____ ____ ____ ____
20-23/____ ____ ____ ____
24-27/____ ____ ____ ____
28-31/____ ____ ____ ____
32-35/____ ____ ____ ____
36-37/____ ____
38-39/____ ____
40-41/____ ____
42-43/____ ____
44-45/____ ____
46-47/____ ____
48-49/____ ____
50-51/____ ____
52-53/____ ____
54-55/____ ____
56-57/____ ____
58-59/____ ____
60-61/____ ____
62-63/____ ____
64-65/____ ____
66-67/____ ____
68-69/____ ____
70-71/____ ____
72-73/____ ____
74-75/____ ____
76-77/____ ____

CARD 5 BEG

1/5
2-5/____ ____ ____ ____
6-7/____ ____
8-9/____ ____
10-11/____ ____
12-13/____ ____
14-15/____ ____
16-17/____ ____
18-19/____ ____
20-21/____ ____
22-23/____ ____
24-25/____ ____
26-27/____ ____
28-29/____ ____
30-31/____ ____
32-33/____ ____
34-35/____ ____
36/____
37/____
38/____
39/____
40/____
41/____
42/____

TURN PAGE FOR REVIEW OF INDIVIDUAL DAYS OF CARE: NEED FOR CONTINUED HOSPITALIZATION

EXPLANATORY NOTES FOR LEVEL OF SERVICE CRITERIA
(THOSE CRITERIA MARKED BY AN ASTERISK)

1. **PROCEDURE PERFORMED:** See manual, Appendix A for list of Class 1 procedures

9. **IV FLUIDS OR MEDS:** Any, except to keep vein open (TKO)

   Includes: Calcium or Potassium Supplements or Antagonists
   Electrolyte Replacement
   Fluids (excludes TKO)
   Hyperalimentation, parenteral
   Meds (Chemotherapy)
   Parenteral Digitalization
   Total Parenteral Nutrition (TPN)

10. **MONITORING PROGRAM IN EFFECT:** Must be monitored more often than once per shift

    Any internal body cavity monitor, such as:

    Cardiac Monitor
    Central Venous Pressure Monitor
    Distal Esophageal Monitor
    Intracranial Pressure Monitor
    Swan Ganz

    Medications (as, monitoring of meds for cardiac or ICU patients)

    Orientation Checks

    Pupil Reaction

    Urine Output

    Vital Signs

11. **RESPIRATORY CARE:** To be used only when there is documentation of continuing need for the service or that improvement is expected to continue with treatment

    Chest PT 3 times daily (by Respiratory Therapist or ICU personnel), plus 1 day

    IPPB 4 times daily, plus 2 days

    Lavage, Bronchial or Tracheal

    Mechanical Respirator

    Mist Tent

    O_2, continuous

    Ventilatory assistance (continuous or standby)

21. **PSYCHIATRIC CARE:** (Includes substance abuse care)

    Electroconvulsive Therapy (ECT), plus 1 day

    Psychotherapy, 3 times weekly

    Restraints (to protect self or others) — does NOT include soft or supportive restraints

    Substance Abuse Program: detoxification or active organized recovery program up to 21 days

    Suicide precautions

26. **MEDICATION ADJUSTMENT OR REGULATION:**

    Antiarrythmics

    Anticoagulation regulation

    Chemotherapeutic agents: Lithium Carbonate, major tranquilizers, tricyclic and monamine oxidase-inhibitor antidepressants

    Diuretic regimen for Ascites

    Insulin therapy regulation (change in dosage more often than every 2 days)

    Psychotropic drugs — 3 or more concurrently

    SubQ heparin therapy

    Ulcer medication regulation (change in dosage more often than every 2 days)

27. **GENERAL OBSERVATION:** 2 days maximum allowed except:

    — Following Acute Myocardial Infarction or transfer from ICU or CCU — 3 days
    — Organ Transplant Rejection — no time limit
    — Patients receiving oral medications — no time limit as long as lab tests remain abnormal
    — Post-op/Post delivery observation: 2 days allowed except for venograms, arteriograms, D&C, biopsies and endoscopies, which are allowed only 1 day

Observation time limits may be extended by specific physician orders for observation of anticipated complications.

REVIEW OF INDIVIDUAL DAYS OF CARE: NEED FOR CONTINUED HOSPITALIZATION

CRITERIA FOR CONTINUED HOSPITALIZATION Check the primary criterion met on each day, then go to Section II	Date:	(43-46)	(47-50)	(51-54)	(55-58)	(59-62)	(63-66)	(67-70)
		(8-9)	(10-11)	(12-13)	(14-15)	(16-17)	(18-19)	(20-21)
OPERATIONS/PROCEDURES:								
1. CLASS 1 PROCEDURE PERFORMED THIS DATE	1.							
2. PRE-OP PREPARATION DAYS: Allow 1 day, except for bowel prep allow 2 days.	2.							
PATIENT CONDITION FACTORS:								
3. CARDIAC (CHEST) PAIN CONTINUES IN PATIENT OFF TELEMETRY 3 DAYS WITH FULL AMBULATION	3.							
4. COMPLICATIONS DEVELOP (Iatrogenic or other)	4.							
5. FOCAL NEUROLOGICAL DEFICIT UNSTABLE (and etiology remains undetermined)	5.							
6. HEMORRHAGE OR PURULENT DRAINAGE, any site, plus 1 day (does not include minimal bleeding, as hemorrhoidal or menstrual)	6.							
7. INABILITY OF PATIENT TO CARE FOR SELF EVEN WITH THE HELP OF ORGANIZED HOME CARE OR FAMILY (does NOT apply to long-standing conditions)	7.							
8. INTAKE/OUTPUT REMAINS ABNORMAL: • Unable to pass flatus/fecal material with regularity, plus 1 day • Unable to void or drain urine (<800 cc/24 hours), plus 1 day • Unable to tolerate prescribed diet or tube feedings (evidenced by nausea or vomiting), plus 2 days	8.							
9. LABORATORY FINDINGS REMAIN OUTSIDE SPECIFIED RANGES	9.							
10. OCULAR PRESSURE (GLAUCOMA) REMAINS >30, PLUS 1 DAY	10.							
11. PAIN CONTINUES DESPITE ANALGESICS OR NARCOTICS, PLUS 2 DAYS	11.							
12. PNEUMONIA PRESENT, PLUS 3 DAYS	12.							
13. SEIZURE ACTIVITY CONTINUES, PLUS 2 DAYS	13.							
14. SKIN CONDITIONS CONTINUE TO REQUIRE INTENSIVE DERMATOLOGICAL CARE, as in Cellulitus or Decubitus showing no improvement	14.							
15. VERTIGO DOCUMENTED, PLUS 1 DAY	15.							
16. VITAL SIGNS REMAIN ABNORMAL, PLUS 1 DAY	16.							
17. WOUND NECROSIS OR INFECTION	17.							
SERVICES CONTINUING:								
18. DRAINS: Still present, or removed, plus 1 day	18.							
19. ICU/CCU CARE STILL IN PROGRESS, PLUS 3 DAYS	19.							
20. ISOLATION CONTINUES	20.							
21. MEDICATION REGULATION CONTINUES (IM or IV) • Anticoagulation therapy (can be oral if pro time remains abnormal) • Steroids, plus 1 day • Other major drugs (insulin, antibiotics, antihypertensives, antiarrythmics, etc.), plus 2 days • 3 psychotropic drugs concurrently (can be oral)	21.							
22. PATIENT REMAINS IN AN ORGANIZED REHABILITATION PROGRAM (e.g., substance abuse, speech therapy) [Except when documentation indicates maximum benefit has been obtained.]	22.							
23. PSYCHIATRIC MANAGEMENT: Allow 3 days past the last documented evidence of improvement, or if evaluation by a psychiatrist is pending.	23.							
24. TELEMETRY CONTINUES	24.							
OBSERVATION CONTINUING:								
25. POST OP/POST DELIVERY OBSERVATION, 2 DAYS (Except 1 day for venograms, arteriograms, D&C, biopsy or endoscopy)	25.							
26. PENDING EKG CONFIRMATION OF CARDIAC DAMAGE (3 days) post occurrence	26.							
DAILY EVALUATION NEED FOR CONTINUED HOSPITALIZATION APPEARS: (Check One Below)		(22-23)	(24-25)	(26-27)	(28-29)	(30-31)	(32-33)	(34-35)
JUSTIFIED — Criterion met	✓ = 31							
JUSTIFIED — No criterion met, but logical extension applies. (Enter criterion # and explain on attached page.)	✓ =							
JUSTIFIED — No criterion met, but record review supports. (Explain on attached page.)	✓ = 32							
NOT JUSTIFIED — No criterion met	✓ = 33							
NOT JUSTIFIED — Criterion met, but record review does not support. (Explain on attached page.)	✓ = 34							
Requires REASON CODE from List 1 → Most Likely (Required) (36-49)								
Second (Optional) (50-63)								
Third (Optional) (64-77)								

CARD 5 CONTD

43-46/____ ____ ____ ____
47-50/____ ____ ____ ____
51-54/____ ____ ____ ____
55-58/____ ____ ____ ____
59-62/____ ____ ____ ____
63-66/____ ____ ____ ____
67-70/____ ____ ____ ____

CARD 6 BEG

1/6
2-5/____ ____ ____ ____
6-7/____ ____
8-9/____ ____
10-11/____ ____
12-13/____ ____
14-15/____ ____
16-17/____ ____
18-19/____ ____
20-21/____ ____
22-23/____ ____
24-25/____ ____
26-27/____ ____
28-29/____ ____
30-31/____ ____
32-33/____ ____
34-35/____ ____
36-37/____ ____
38-39/____ ____
40-41/____ ____
42-43/____ ____
44-45/____ ____
46-47/____ ____
48-49/____ ____
50-51/____ ____
52-53/____ ____
54-55/____ ____
56-57/____ ____
58-59/____ ____
60-61/____ ____
62-63/____ ____
64-65/____ ____
66-67/____ ____
68-69/____ ____
70-71/____ ____
72-73/____ ____
74-75/____ ____
76-77/____ ____

TURN PAGE TO RECORD NOTES, IF NEEDED

EXPLANATORY NOTES FOR NEED FOR CONTINUED HOSPITALIZATION

* **7. INABILITY OF PATIENT TO CARE FOR SELF EVEN WITH THE HELP OF ORGANIZED HOME CARE OR FAMILY:**

(Does not apply to longstanding conditions)

- Unable to ambulate independently or with walker, cane, crutches, wheelchair or prosthesis
- Unable to dress/clean/care for feeding tube, stoma, appliance, salivary cutaneous fistula, surgery wounds, drainage tubes (includes tracheostomy, urinary drain)
- Unable to perform Activities of Daily Living (ADL) due to wheezing, dyspnea, cyanosis or physical limitation

* **9. LABORATORY FINDINGS REMAIN OUTSIDE SPECIFIED RANGES:**

- Blood sugar: <75 or >105; fasting, <60 or >90; plus 2 days
- Blood calcium: <9 or >11 (ionized calcium <4.5 or >5.6); plus 2 days
- BUN: >35; plus 4 days
- Creatinine: >2, plus 4 days
- Digitalis blood level: <1 or >2 units/ml
- Prothrombin time: <11 or >16 seconds (or indicated as abnormal by % of control); plus 2 days
- WBC: <2500 (patients on chemotherapy) plus 3 days
- Other Lab values remain outside normal range

* **12. PNEUMONIA**

Any bacterial pneumonia still present, except pneumococcal, or,
Pneumococcal or other type of pneumonia present with one of the following:

Age very young (<5) or very old (>65)

General debilitated condition of patient

Presence of Diabetes Mellitus

* **16. VITAL SIGNS REMAIN ABNORMAL, PLUS 1 DAY**

Temperature: >99° after discontinuation of antipyretics
>100° with neutrophil count <2000
>102° with bacteria by culture/smear or with WBC >12,000

Pulse: <50 or >140

Respiratory Rate: >32

Blood Pressure: Systolic <80 or >200
Diastolic >120

REASON CODES

List 1: ADMISSION OR CONTINUED HOSPITALIZATION APPEARS NOT JUSTIFIED

PATIENT REQUIRES NO INSTITUTIONAL CARE, BUT IS ADMITTED OR RETAINED BECAUSE:

1. Terminal patient for humanitarian reasons
2. Patient, family, physician prefer/insist on acute care. Includes non-compliant patient (*refuses* to comply)
3. Delay in orders or arrangements for discharge
4. Delay in discharge pending completion of patient or family education for home care
5. No home or alternate care is possible or available. (Includes patients who *cannot* comply, hostile family situation, or adverse home or environmental factors)
6. Other. Explain in Notes.
7. No reason can be deduced from available information

PATIENT REQUIRES SUBACUTE LEVEL OF CARE, BUT IS ADMITTED OR RETAINED BECAUSE:

21. Terminal patient for humanitarian reasons
22. Condition improving, recovery anticipated within 3 days
23. No lower level of care available: facilities will not accept patient or no beds available
24. Difficulty in placing patient in lower level facility. Includes Financial and Medical Difficulties, as:
    - Awaiting medical clearance from alternative facility
    - Awaiting financial clearance from alternative facility
    - Insurance (or Medicare/Medicaid) coverage is broader in an acute care setting, or 3 day Medicare qualification is needed
25. Difficulty in arranging transfer to other facility or delayed initiation of discharge planning
26. Patient, family or physician prefers or insists on acute level hospitalization. Includes non-compliant patient, teaching or research reasons, and legal reasons (controversial patient; psychiatric requirements; real, implied or feared threat of malpractice).
27. Other. Explain in Notes.
28. No reason can be deduced from available information

List 2: LEVEL OF SERVICE APPEARS TO BE NONACUTE

DELAYS IN SCHEDULED SERVICE DUE TO:

31. Tests or consultations not complete or results not yet available
32. Change in patient condition
33. Patient or family noncompliant or indecisive regarding treatment course
34. Resource availability precludes timely delivery of service. Includes both physician and hospital causes, as, hours of operation (40 hour work week), staff shortages, emergency situations which may "bump" patient, or scheduling conflicts or delays, including transfers to other facilities.
35. Sequencing conflicts or delayed ordering by hospital or physician. Includes premature admission (e.g., Friday admission to hold bed for Monday surgery, sequencing of tests causes extended stay, etc.).
40. Patient does not require acute level services (i.e. patient did not need to be admitted or patient is ready for discharge).
50. Other. Explain in Notes.
60. No reason can be deduced from available information.

PHASE OF STAY

PRE-TREATMENT PHASES:

1 Diagnostic work-up in progress (prior to determination of definitive therapeutic course).
NOTE: Phase 1 applies only when changes in the patient condition or diagnosis indicates the patient is in an "evolving" status. Once a therapeutic course has been determined, use Phase 2 if no therapy has begun, or Phase 3, 4, or 5 if therapy is underway.

2 Scheduled procedure, treatment, or management pending (prior to initiation of definitive therapeutic course).
NOTE. Phase 2 may be used for patients entering hospital whose treatment course has been determined prior to admission.

TREATMENT PHASES:

3 Surgical — includes day of procedure and post-op observation period; as defined in Level of Service criteria.

4 Medical Therapy — includes drug, respiratory or physical therapy, psychotherapy, or traction.

5 Transition from one therapeutic regimen to another: Use when any therapy is being given, but focus is changing. If significant complications develop, or entire course of treatment is being re-evaluated use Phase 1.

POST-TREATMENT PHASE:

6 Planned course of therapy completed

ATTACHMENT C

SURGERY APPROPRIATENESS EVALUATION PROTOCOL

PATIENT'S NAME _____ FORM CONTROL # _____
(LAST) (FIRST) (INITIAL)

HOSPITAL NAME _____ PATIENT'S RECORD # _____

SURGERY APPROPRIATENESS EVALUATION PROTOCOL CARD 1 1/1

Form Control # _____ Year ____ 2-6/ 7-8/
(2-6) (7) (8)

Hospital Name _____ Hospital Code # _____ 9-11/
(9-11)

Patient's Age (Last Birthday) ____ ____ Patient's Sex (check) (14) M ( )1 F ( )2 12-13/ 14/
(12) (13)

Primary Insurance Coverage (check one only): (15)
Medicare ( )1
Medicaid ( )2
Blue Cross ( )3
Commercial ( )4
Other Third Party ( )5
Self-Pay ( )6

Hospital Service on Admission (check): (16)
Pediatrics ( )1
Medicine ( )2
Surgery ( )3
Gynecology ( )4
Psychiatry ( )5
Obstetrics ( )6
Other ( )7

Hospital Service on Date Being Reviewed (check): (17)
Pediatrics ( )1
Medicine ( )2
Surgery ( )3
Gynecology ( )4
15/ 16/ 17/

Admission Date |__|__|/|__|__|/__|__| 18-22/
(18-22) (Month) (Day) (Year)

Date Being Reviewed |__|__|/|__|__|/__|__| 23-27/
(23-27) (Month) (Day) (Year)

Length of Stay (days) ___ 28-30/
(28-30)

II. Diagnoses, active, this admission:

A. _____ ICD-9-CM |__|__|__.__|__| 31-35/

B. _____ ICD-9-CM |__|__|__.__|__| 36-40/

C. _____ ICD-9-CM |__|__|__.__|__| 41-45/

III. Major Procedures, this admission, on or before day reviewed:

A. Therapeutic Date 46-49/

1. _____ ICD-9-CM |__|__|.__|__| 50-53/
Month/Day

2. _____ ICD-9-CM |__|__|.__|__| 54-57/
Month/Day 58-61/

B. Diagnostic

1. _____ ICD-9-CM |__|__|.__|__| 62-65/
Month/Day 66-69/

2. _____ ICD-9-CM |__|__|.__|__| 70-73/
Month/Day

IV. Reason for Admission (optional): _____ 74-77/

_____ 78-79/

CARD 2 1/2

2-6/

Criteria of Admission

LOCATION OF SURGERY/RISK FACTOR ASSESSMENT

1. Comorbidity

A. Respiratory Status

	NO	YES	
1. Significantly abnormal pulmonary function measurements:			
a. FVC < 1.0 L	(0) ___	(1) ___	7/
b. FEV_1/FVC < 50%	(0) ___	(1) ___	8/
c. Arterial pCO_2 > 50 mmHg (on room air)	(0) ___	(1) ___	9/
d. Arterial pO_2 < 50 mmHg (on room air)	(0) ___	(1) ___	10/
2. Sleep apnea	(0) ___	(1) ___	11/

B. Significant Co-Morbid Disease (must be documented)

1. Blood disorders			
a. SS or SC disease	(0) ___	(1) ___	12/
b. Hemophilia	(0) ___	(1) ___	13/
c. Idiopathic thrombocytopenic purpura	(0) ___	(1) ___	14/
2. Cardiac diseases			
a. Angina pectoris Class III or IV (NYHA) currently	(0) ___	(1) ___	15/
b. Congestive heart failure Class III or IV (NYHA) currently	(0) ___	(1) ___	16/
c. Myocardial infarction within 90 days	(0) ___	(1) ___	17/
3. Personal or family history of malignant hyperthermia	(0) ___	(1) ___	18/
4. Patients with documented difficulty regulating medications for:			
a. Endocrine disease (diabetes, Addison's disease, thyrotoxicosis)	(0) ___	(1) ___	19/
b. Hypertension	(0) ___	(1) ___	20/
c. Bronchospastic lung disease	(0) ___	(1) ___	21/
d. Seizures	(0) ___	(1) ___	22/

2. Potential for Complications

A. Type of Procedure

	NO	YES	
1. Surgery on an internal organ, including procedures on head, neck, and back, as well as on thoracic, abdominal, and pelvic organs	(0) ___	(1) ___	23/
2. Blind biopsy of an internal organ	(0) ___	(1) ___	24/
B. General or regional anesthesia lasting more than 90 minutes	(0) ___	(1) ___	25/
Document actual anesthesia time: _____			26-28/
C. Social factors precluding prompt access to medical attention, in case of adverse post-procedure effect			
1. Lack of ability to communicate, because of living alone or telephone inaccessibility	(0) ___	(1) ___	29/
2. Lack of practical transportation availability; great distance from urgent medical care	(0) ___	(1) ___	30/
3. Mental instability	(0) ___	(1) ___	31/
Specify condition: _____			32/

3. Need for Intensive Post-Operative Care

A. Amputations, except digits	(0) ___	(1) ___	33/
B. Peripheral vascular surgery	(0) ___	(1) ___	34/
C. Placement of orthopedic hardware, except distal K-wire insertion for stabilization	(0) ___	(1) ___	35/
D. Placement of drainage tubes	(0) ___	(1) ___	36/

OVERRIDE OPTIONS

Other appropriate services or conditions:

Yes 1( ) Description: _____ 37/
37 _____
_____ 38-39/

Criteria met, but inappropriate nevertheless:

Yes 1( ) Description: _____ 40/
40 _____
_____ 41-42/

SURGERY APPROPRIATE ON AN INPATIENT BASIS (0) ___ (1) ___ 43/

TIMELINESS OF ADMISSION

	NO	YES	
A. Surgery or procedure performed within 24 hours of admission (inpatient)	(0) ___	(1) ___	44/

IF NO, COMPLETE THE FOLLOWING:

Justifications for Not Doing Surgery the Day of or the Day After Admission

B. Special pre-operative evaluation/treatment, available only on an inpatient basis			
1. Supervised diet	(0) ___	(1) ___	45/
2. Parenteral medications	(0) ___	(1) ___	46/
3. Extensive enemas (more than a Fleets enema)	(0) ___	(1) ___	47/
4. Procedures such as angiography, endoscopy, myelography, not to be done as part of the planned surgery	(0) ___	(1) ___	48/
5. Dialysis or exchange transfusions	(0) ___	(1) ___	49/
C. Patient Condition			
1. Unacceptable cardiac status			
a. Suspicion of ongoing or recent myocardial infarction	(0) ___	(1) ___	50/
b. Uncontrolled or unstable angina pectoris	(0) ___	(1) ___	51/
c. New or complex arrhythmia	(0) ___	(1) ___	52/
d. Uncompensated congestive heart failure	(0) ___	(1) ___	53/
2. Unacceptable cerebrovascular status			
a. New stroke not completed	(0) ___	(1) ___	54/
b. Transient ischemic attacks	(0) ___	(1) ___	55/
3. Unacceptable pulmonary status			
a. Unrelieved bronchospasm	(0) ___	(1) ___	56/
b. Documented deterioration of chronic obstructive lung disease	(0) ___	(1) ___	57/
4. Unacceptable hematologic status			
a. Unexpected anemia requiring transfusion or explanation pre-operatively	(0) ___	(1) ___	58/
b. New granulocytopenia ($<1500/mm^3$) or thrombocytopenia ($<100,000/mm^3$) requiring explanation pre-operatively	(0) ___	(1) ___	59/
c. Severe thrombocytopenia or lack of other clotting factors (e.g., prothrombin) not correctable in time (<24 hours)	(0) ___	(1) ___	60/

		NO	YES	
5.	Unacceptable metabolic status			
	a. Uncontrolled diabetes mellitus	(0) ___	(1) ___	61/
	b. Severe (Cr>5.0 mg/dl) or new azotemia	(0) ___	(1) ___	62/
	c. Severe liver dysfunction, other than clotting (transaminases 5x upper limit of laboratory normal)	(0) ___	(1) ___	63/
	d. Uncontrolled hyperthyroidism or uncorrected hypothyroidism	(0) ___	(1) ___	64/
	e. Uncorrected electrolyte disturbances Sodium, potassium, or calcium outside hospital's own laboratory normal ranges	(0) ___	(1) ___	65/
6.	Unacceptable mental status			
	a. New confusion or coma	(0) ___	(1) ___	66/
	b. Incompetence or inability to understand operative permit, etc.	(0) ___	(1) ___	67/
7.	Uncontrolled seizures	(0) ___	(1) ___	68/
8.	Unexplained new rash	(0) ___	(1) ___	69/
9.	Active infection, other than that for which surgery is planned	(0) ___	(1) ___	70/
10.	Unexplained fever, if not related to need for surgery	(0) ___	(1) ___	71/
D.	Cancellation of surgery because of unforeseen administrative/technical circumstances	(0) ___	(1) ___	72/

OVERRIDE OPTIONS

Other appropriate services or conditions:

Yes 1( ) Description: _____ 73/
73 _____
_____ 74-75/

Criteria met, but inappropriate nevertheless:

Yes 1( ) Description: _____ 76/
76 _____
_____ 77-78/

ADMISSION APPROPRIATE ON THIS DAY (0) ___ (1) ___ 79/

ATTACHMENT D

TECHNICAL NOTE

WEIGHTED RELIABILITY COEFFICIENTS

TECHNICAL NOTE*

WEIGHTED RELIABILITY COEFFICIENTS

As discussed previously, the sample of cases chosen for the reliability trials (N = 99) was (intentionally) biased toward admissions found to be non-acute by the AEP. This biased reliability trial sample may, in part, account for the observed discrepancies in reliability coefficients and kappa statistics between our research and that reported by the (SMI) instrument developers. The reliability of these criteria, as might be found to occur in a random sample of cases, where acute admissions occur more frequently, is of general and theoretical importance. A weighted analysis of reliability answers the question: What is the expected reliability of the AEP and the SMI if the reliability coefficients are based on a random, unbiased sample of cases?

Methods for Weighted Reliability Analysis

The analysis is conducted by weighting the data from the reliability trials so the distribution of cases in the four cells defined by the cross tabulation of (the initial administration of) the AEP and the SMI is the same as the distribution in the random population sample of 1,173 cases. The reliability statistics are then calculated on the weighted data. See Table TN - 1 for the distribution of the reliability trial sample and Table TN - 2 for the 2 x 2 distribution of the random population sample used to derive estimates of non-acute care.

---

* Analysis on weighted reliability was completed with consultation from Graham Kalton, Ph.D., Director, Sampling Section, Institute for Social Research, The University of Michigan, Ann Arbor, Michigan.

Of the 99 cases chosen for the reliability trials, 94 (94.95%) were reviewed by two (2) different RNs applying the AEP. As indicated in Table TN - 3, the overall AEP reliability is 78.72% (74/94) (kappa = .46, $p < .01$), while the specific non-acute reliability is 74.68% (59/79) and the acute reliability is 42.86% (15/35).

Similarly, as illustrated in Table TN - 4, of the 99 cases in the reliability trial, two (2) different SMI RNs reviewed 98 (98.99%) of the 99 cases. The overall SMI reliability is 73.47% (82/98) (kappa = .03), while the non-acute reliability is 10.34% (3/29) and the acute reliability is 72.63% (69/95).

As discussed in the reliability section, while the overall reliability coefficients of the AEP (78.72%) and the SMI (73.47%) are quite similar, the kappa statistic is significant ($p < .01$) for only the AEP admission reviews. This indicates that the overall agreement for the SMI does not exceed chance association. Basically, this is due to the differences in the non-acute reliability for the AEP (74.68%) and the SMI (10.34%). The SMI, while reliably identifying acute care, does not reliably identify non-acute admissions.

We now turn our analysis to an examination of the within cell reliability for both the AEP and the SMI. These within cell reliabilities are measures of test/retest agreement calculated within cells defined by the original reviews on the SMI and AEP. Because they are calculated within cells, their meanings are somewhat different from the reliability coefficients presented elsewhere in this report. Despite their differences from other reliability coefficients, they are useful, descriptively. More importantly, these within cell reliability coefficients are calculated to compute the weighted estimates of

reliability. To determine if and/or how the process of purposely biasing the reliability sample toward cases found to be non-acute by the AEP affected the overall, specific non-acute, specific acute reliability and kappa statistics, we examine the degree of reliability for the AEP and the SMI within each cell of the 2 x 2 (AEP by SMI) table (Table TN - 3). The analysis demonstrates whether and how the reliability sampling procedure, biased toward cases determined by the AEP to be non-acute, differentially affects the reliability of the AEP and the SMI.

Within Cell Analysis

This analysis examines, in turn, the "within cell reliability" of the AEP and the SMI criteria-sets according to the scheme presented in Table TN - 5.

Table TN - 6 gives the overall within cell reliability for the AEP based on the cases which both the AEP and SMI agree were acute (Cell 1). The data show that the overall AEP reliability for cases identified by both the AEP and SMI to be acute for the AEP is 87.5% (14/16). Similarly, as indicated in Table 7, the overall within cell reliability for the SMI for all admissions found to be acute by both the AEP and SMI is 88.24% (15/17) (see Table TN - 7).

These data demonstrate that within cell reliability for both the AEP (87.5%) and SMI (88.24%) for all cases identified by both the AEP and SMI to be acute (Cell 1), which accounts for about 17% of the reliability sample, is virtually identical. Both instruments are equally reliable on cases both criteria-sets indicate as acute (e.g., severely ill patients in need of intense, acute hospital, inpatient services).

However, we discover differences between the AEP and the SMI when we examine the reliability of both criteria-sets on cases both the AEP and SMI categorize as non-acute (Cell 4). The reliability of the AEP for these cases is 81.82% (9/11), while the reliability of the SMI for these cases is 27.27% (3/11) (see Tables TN - 8 and TN - 9). The AEP reliably identifies these cases as non-acute while the SMI does not reliably identify such admissions as non-acute. The weakness of the SMI appears to be in reliably identifying non-acute admissions.

Turning our attention to the subset of cases in the reliability trials found to be acute by the AEP and non-acute by the SMI (Cell 3), we find that cases seldom fell into this category since it is extremely unlikely for the AEP to find care acute and the SMI to find the same admission to be non-acute. The data indicate that the within cell reliability of the AEP for these unique and unusual cases is higher for the AEP (33.33%) than for the SMI (0%) (see Tables TN - 10 and TN - 11).

Finally, we turn our attention to a critical comparison: The majority of the cases in the reliability trials identified to be non-acute by the AEP and acute by the SMI (Cell 2), representing 68.7% of the reliability sample. These cases constitute the "grey" cases in which the two criteria-sets disagree on the evaluation of the appropriateness of hospitalization. These cases represent cases in which the severity of illness and need for intense services may be unclear. They also represent a significant oversample of the actual cases in which the two instruments disagree in comparison to the random sample of cases in Cell 2 (28.64%, see Table TN - 2).

The data demonstrate that the overall within cell reliability for both criteria-sets is similar. It is 78.13% (50/64) for the AEP and 80.60% (54/67) for the SMI (see Tables TN - 12 and TN - 13). The critical point to keep in mind in this comparison is that the AEP reliably identifies specific non-acute care, a more difficult task than reliably identifying acute care. In summary (see Table TN - 14), the AEP is uniformly reliable across all conditions (Cells) except in those few instances where the cases are categorized by the AEP to be acute and categorized by the SMI to be non-acute.

As is evident from Table TN - 14, the SMI is reliable in instances in which the SMI identifies care as acute. The reliability of the SMI is low in instances in which the SMI originally identified care to be non-acute. These within cell analyses support the reliability finding presented in the chapter on instrument reliability. We now turn our attention to the weighted reliability coefficients to determine the expected reliability of these instruments had they been applied to a randomly selected and unbiased sample of admissions (e.g., Table TN - 2).

Appropriateness Evaluation Protocol

It can be argued that the reliability coefficients presented in this research do not reflect actual reliability coefficients for the two criteria-sets found in a random sample of cases. Our reliability trial sample is biased toward non-acute care (68% of the cases form Cell 3 in the reliability sample, while 28.64% of the cases form Cell 3 in the random sample from a normal population). The unweighted reliability trial sample contains more non-acute care, and proportionally more disagreement between the AEP and the SMI than would be found in a random sample of cases. To adjust for these sampling differences,

weighted reliability coefficients are provided for each of the two criteria-sets. These weighted reliability coefficients reflect the relationship of the reliability trial sample to the population sample of agreements/disagreements between the AEP and SMI for within cell admissions.

Table TN - 15 displays the weighted reliability of the AEP based on our probability sample and population sample fractions. In Cell 1, where both the instruments agree (on the original review) that care is acute, there are 758 cases in the population (see Table TN - 2). From Table TN - 6, the fraction of cases that the AEP again found acute on re-review is 0.875 (14/16). Thus, 663.25 cases (758 x 0.875) are classified as belonging to Cell 1. For Cell 2, the distribution fraction is (2/16). Applying this distribution fraction to the 758 cases, gives us 94.75 cases as the number (of these 758) that we would expect to find in Cell 2. Since the distribution fractions for Cell 3 and Cell 4 are both zero, none of the 758 cases are distributed to these cells in the weighted population estimates.

In Cell 3, where the AEP indicates (on first application) that care is acute while the SMI indicates the care is non-acute, there are 15 cases in the population (Table TN - 2). As indicated in Table TN - 10, the distribution fraction of 1/3 from the reliability sample applied to the 15 cases from the population results in 5 cases distributed to Cell 1. For Cell 2, the distribution fraction is 2/3, thus 10 cases are included in this cell for the weighted population estimates. Since the distribution fractions for Cells 3 and 4 are zero, none of the 15 cases is distributed between these cells.

In Cell 2 for the population distribution, where the SMI finds acute care, while the AEP finds non-acute care [on first review, there are 336 cases in the population (Table TN - 2)]. As indicated in Table TN - 12, the distribution fraction of 14/64 results in the distribution of 73.5 cases into Cell 3 in the weighted reliability estimates. Cell 4 receives 262.5 cases (336 x 50/64). Cells 1 and 2 receive none of the 336 cases since their distribution factors are zero. Cell 4, where both the AEP and the SMI agree (on initial review) that care is non-acute, has 64 cases in the population (Table TN - 2). From Table TN - 8, the distribution of these cases to Cell 4 is 52.36 (64 x 9/11) while Cell 3 receives 11.64 cases (64 x 2/11). Cells 1 and 2 receive none of the distribution of these cases.

As indicated in Table TN - 16, the weighted reliability coefficients, based on the reliability sample and population sample fractions, indicate that the overall expected reliability of the AEP for the random population sample is 83.81% (kappa = .64, $p < .01$) while the expected non-acute reliability coefficient is 62.38%. (Note that the specific non-acute reliability for the AEP drops from 74.68% to 62.38% when a random non-biased sample is estimated, however, the kappa statistic increases from .47 to .64 in the weighted analysis.)

Standardized Medreview Instrument

A similar weighting scheme applies to the reliability trials based on the population categorization for the SMI (see Table TN - 17). Results show an overall weighted reliability of 81.59% (kappa = .05), a specific non-acute reliability coefficient of 7.48%, and a specific acute reliability of 81.31% (see Table TN - 18). These data (which are weighted to estimate the results that would be expected from a random sample) confirm the low specific of SMI reliability. There is only minimal change in the SMI reliability coefficients in comparing the weighted estimates of the population reliability coefficients with the actual unweighted reliability coefficients.

These data demonstrate that the SMI, when applied to a random sample of cases (N = 1,173) will be found to be unreliable for admission reviews primarily because of its low specific non-acute reliability coefficient. As indicated in Table TN - 19, when adjusting the reliability coefficient to adjust for the sampling procedure, the reliability (kappa statistic) of the AEP improves, while there is little change in the reliability of the SMI. These data confirm our earlier findings supporting the reliability of the AEP.

TABLE TN - 1

Inter-Rater Admission Reliability Sample (N = 99)

		AEP		
		Acute	Non-Acute	Total
SMI	Acute	17 (17.2%)	68 (68.7%)	85 (85.9%)
	Non-Acute	3 (3.0%)	11 (11.1%)	14 (14.1%)
	Total	20 (20.2%)	79 (79.8%)	99 (100%)

TABLE TN - 2

Inter-Rater Admission Reliability Population (N = 1,173)

		AEP		
		Acute	Non-Acute	Total
SMI	Acute	758 (64.62%)	336 (28.64%)	1,094 (93.26%)
	Non-Acute	15 (1.28%)	64 (5.46%)	79 (6.74%)
	Total	773 (65.9%)	400 (34.1%)	1,173 (100%)

TABLE TN - 3

AEP Inter-Rater Admission Reliability Trials ($\underline{N}$ = 94)

		AEP Re-Review		
		Acute	Non-Acute	Total
AEP Original Review	Acute	15 (15.96%)	4 (4.26%)	19 (20.21%)
	Non-Acute	16 (17.02%)	59 (62.77%)	75 (79.79%)
	Total	31 (32.98%)	63 (67.02%)	94 (100%)

TABLE TN - 4

SMI Inter-Rater Admission Reliability Trials (N = 98)

		SMI Re-Reviews		
		Acute	Non-Acute	Total
SMI Original Review	Acute	69 (70.41%)	15 (15.31%)	84 (85.72%)
	Non-Acute	11 (11.22%)	3 (3.06%)	14 (79.79%)
	Total	80 (32.98%)	18 (67.02%)	98 (100%)

TABLE TN - 5

Inter-Rater Admission Reliability Sample ($\underline{N}$ = 99)

		AEP		
		Acute	Non-Acute	Total
SMI	Acute	Cell 1 Acute Agreement	Cell 2 Disagreement	SMI Acute
	Non-Acute	Cell 3 Disagreement	Cell 4 Non-acute Agreement	SMI Non-Acute
	Total	AEP Acute	AEP Non-Acute	All Cases

TABLE TN - 6

AEP Within Cell Reliability For Admissions Categorized As Acute
By Both AEP and SMI (Cell 1)*

		AEP Re-Review (Inter-rater)		
		Acute	Non-Acute	Total
Original AEP Review	Acute	14 (87.5%)	2 (12.5%)	16 (100%)
	Non-Acute	0	0	0
	Total	14 (87.5%)	2 (12.5%)	16 (100%)

* One of the 17 cases in the sample was not reviewed by a second RN applying the AEP.

TABLE TN - 7

SMI Within Cell Reliability for Admissions Categorized As Acute By Both The AEP and SMI (Cell 1)

		SMI Re-Review (Inter-rater)		
		Acute	Non-Acute	Total
Original SMI Review	Acute	15 (88.24%)	2 (11.76%)	17 (100%)
	Non-Acute	0	0	0
	Total	15 (88.24%)	2 (11.76%)	17 (100%)

TABLE TN - 8

AEP Within Cell Reliability for Cases Categorized As Non-Acute
By Both the AEP and SMI (Cell 4)

		AEP Re-Review (Inter-rater)		
		Acute	Non-Acute	Total
Original AEP Review	Acute	0	0	0
	Non-Acute	2 (18.18%)	9 (81.82%)	11 (100%)
	Total	2 (18.18%)	9 (81.82%)	11 (100%)

TABLE TN - 9

SMI Within Cell Reliability for Cases Categorized As Non-Acute
By Both the AEP and SMI (Cell 4)

		SMI Re-Review (Inter-rater)		
		Acute	Non-Acute	Total
Original SMI Review	Acute	0	0	0
	Non-Acute	8 (72.73%)	3 (27.27%)	11 (100%)
	Total	8 (72.73%)	3 (27.27%)	11 (100%)

TABLE TN - 10

AEP Within Cell Reliability for Cases Categorized As Acute by the AEP and Non-Acute By the SMI (Cell 3)

		AEP Re-Review (Inter-rater)		
		Acute	Non-Acute	Total
Original AEP Review	Acute	1 (33.33%)	2 (66.67%)	3 (100%)
	Non-Acute	0	0	0
	Total	1 (33.33%)	2 (66.67%)	3 (100%)

TABLE TN - 11

SMI Within Cell Reliability for Cases Categorized Acute By the AEP and Non-Acute By the SMI (Cell 3)

		SMI Re-Review (Inter-rater)		
		Acute	Non-Acute	Total
Original SMI Review	Acute	0	0	0
	Non-Acute	3 (100%)	0	3 (100%)
	Total	3 (100%)	0	3 (100%)

TABLE TN - 12

AEP Within Cell Reliability for Admissions Categorized As Non-Acute by the AEP and Acute By the SMI (Cell 2)

		AEP Re-Review (Inter-rater)		
		Acute	Non-Acute	Total
Original AEP Review	Acute	0	0	0
	Non-Acute	14 (21.87%)	50 (78.13%)	64 (100%)
	Total	14 (21.87%)	50 (78.13%)	64 (100%)

TABLE TN - 13

SMI Within Cell Reliability for Admissions Categorized As Non-Acute by the AEP and Acute By the SMI (Cell 2)

		SMI Re-Review (Inter-rater)		
		Acute	Non-Acute	Total
Original SMI Review	Acute	54 (80.60%)	13 (19.40%)	67 (100%)
	Non-Acute	0	0	0
	Total	54 (80.60%)	13 (19.40%)	67 (100%)

TABLE TN - 14

Summary of the AEP and SMI Within Cell Reliability Across All Categories of Agreement/Disagreement

Criteria-Set	AEP/SMI ACUTE	AEP/SMI NON-ACUTE	AEP ACUTE SMI NON-ACUTE	AEP NON-ACUTE SMI ACUTE
AEP	87.50	81.82	33.33	78.13
SMI	88.24	27.27	0	80.60

TABLE TN - 15

The Weighted Reliability of the AEP Based On
Probability Sample and Population Sample Fractions*

		AEP Re-Review (Inter-rater)		
		Acute	Non-Acute	Total
ORIGINAL AEP REVIEW	Acute	CELL 1 663.25 (T6) 5.00 (T10) 668.25	CELL 1 94.75 (T6) 10.00 (T10) 104.75	773 (66%)
	Non-Acute	CELL 3 73.58 (T12) 11.64 (T8) 85.22	CELL 4 262.42 (T12) 52.36 (T8) 314.78	400 (34%)
	Total	753 (64%)	420 (36%)	1,173 (100%)

* See Table 2 for population sample fractions.
(T) = table from which probability sample fraction is derived.

TABLE TN - 16

Comparative Weighted and Unweighted Inter-Rater Reliability Coefficients For the AEP

	Reliability Coefficients			
	OVERALL	SPECIFIC NON-ACUTE	SPECIFIC ACUTE	KAPPA
Weighted	83.81%	62.38%	77.87%	.64*
Unweighted	78.72%	74.68%	42.86%	.47*

* Probability significant $p < .01$.

TABLE TN - 17

The Weighted Reliability of the SMI Based On
Probability Sample and Population Sample Fractions*

		SMI Re-Review (Inter-rater)		
		Acute	Non-Acute	Total
ORIGINAL SMI REVIEW	Acute	CELL 1 668.82 (T7) 270.82 (T13) 939.64	CELL 1 89.18 (T7) 65.18 (T13) 154.36	1,094 (93%)
	Non-Acute	CELL 3 15.00 (T11) 46.53 (T9) 61.53	CELL 4 0.00 (T11) 17.47 (T9) 17.47	79 (7%)
	Total	1,001 (85%)	172 (15%)	1,173 (100%)

* See Table 2 for population sample fractions.
(T) = table from which probability sample fraction is derived.

TABLE TN - 18

Weighted and Unweighted Inter-Rater Reliability Coefficients For the SMI

	Reliability Coefficients			
	OVERALL	SPECIFIC NON-ACUTE	SPECIFIC ACUTE	KAPPA
Weighted	81.59%	7.48%	81.31%	.05
Unweighted	73.47%	10.34%	72.63%	.03

TABLE TN - 19

Expected Weighted Reliability Coefficients

	OVERALL	SPECIFIC NON-ACUTE	SPECIFIC ACUTE	KAPPA
AEP	83.81%	62.38%	77.87%	.64*
SMI	81.59%	7.48%	81.31%	.05

* Probability significant $p < .01$.

ATTACHMENT E

Addendum

Reliability Of The Standardized Medreview Instrument: Re-Analysis Using PRO Reviewers

Standardized Medreview Instrument Reliability

Since the reliability coefficients for the SMI were low in relation to those reported by the instrument developer (SysteMetrics), additional reliability reviews were conducted employing a different group of raters from a PRO which uses the SMI. Thus, an attempt to confirm previous results with raters who had extensive experience (2 years) using the Standardized Medreview Instrument was conducted.

In an effort to duplicate field conditions, we contacted a PRO in Indiana (Peerview) and arranged for two RNs, each with at least two years of experience with the SMI, to conduct additional reliability reviews. These reviews were conducted the week of November 3, 1986. The Peerview auditors were given copies of the same charts (N = 99) used in the reliability section of our research project (also reviewed by a representative of SysteMetrics, Ms. Kathy Barnes, A.R.T.). In total, over the course of one week, Peerview auditors reviewed the charts which corresponded exactly to the 99 charts used in the reliability sample of our research. Our analysis entails a three-way comparison with the 99 charts reviewed by Kathy Barnes compared to the same 99 charts reviewed by each of the PRO reviewers and by BCBSM reviewers. Only admission reviews were performed.

Since the initial AEP and SMI reliability reviews were performed by a group of BCBSM nurses newly trained in the use of the SMI, the subsequent SMI reviews by PRO reviewers should, to some extent, result in higher reliability coefficients. The question is: How much higher are the subsequent SMI agreement

coefficients? These reviews should represent the absolute maximum reliability of the SMI under ideal field conditions. However, it should be noted that a straightforward comparison between the reliability of the AEP with the reliability of the SMI employing experienced raters is, in some sense, an unfair comparison. The reliability coefficients for the AEP were obtained with newly trained auditors. We might expect an increase in AEP reliability coefficients were we to use AEP reviewers who had several years of experience with the criteria set. Nevertheless, given the importance of these criteria in utilization review activities, and the low reliability coefficients found with earlier application of the SMI, we find it important to conduct these additional reviews.

Results

The additional reliability trials conducted in cooperation with Peerview resulted in improved reliability coefficients with the SMI. While these reliability coefficients are similar to those recently reported by Syste-Metrics (Moynihan, personal communication, dated August 13, 1986), under their SuperPRO contract with HCFA, the SMI reliability coefficients (36% - 45%) are nevertheless below those obtained with newly trained RNs using the AEP (75%). The inter-rater reliability coefficients obtained by PRO reviewers (36% - 45%) are similar to the intra-rater reliability coefficients obtained by BCBSM RNs newly trained in the use of the SMI (29%).

This suggests that the low SMI inter-rater reliability coefficients obtained with BCBSM RNs newly trained in the use of the SMI is due to the lack of consistency between raters. The improved (but lower than AEP) reliability coefficients obtained by the PRO reviewers suggests that extensive

experience within the same organizational setting improves consistency in applying the SMI criteria. It is also possible that an improvement in reliability coefficients would be found if AEP reviewers from the same organization with several years of application experience were used.

The non-acute reliability coefficient between the SysteMetrics (hereafter referred to as the SuperPro) reviewer, Ms. Kathy Barnes, A.R.T., and PRO reviewer 1 is 45%; the non-acute reliability coefficient between the SuperPRO reviewer and the PRO reviewer 2 is 36%, while the non-acute reliability coefficient between PRO reviewer 1 and PRO reviewer 2 is 43% (see Table A - 1, attached). While the kappa statistic for the overall reliability coefficient for each of these three (admission reliability) comparisons is statistically significant ($p < .01$), the inter-rater admission reliability coefficients and kappa statistic (which adjusts for chance agreement and differences in reliability due to differences in the base rates of non-acute care) for these reviews are below those obtained using the AEP (75%).

Thus, use of an identical set of cases to compare earlier SMI reviews with SMI reviews performed by experienced PRO and SuperPRO reviewers, shows that use of the SMI by experienced PRO reviewers results in statistically significant but lower rates of admission reliability in comparison to the AEP. Our earlier conclusion, based on reviews performed by BCBSM RNs, that the SMI is less reliable than the AEP appears to be supported.

Since there is some question about the effect which the use of overrides has on reliability, we performed an analysis on the same set of cases, using the SMI but disregarding evaluation changes due to the use of an override (thus all evaluations are based on "objective" criteria rather than "subjective" overrides). These data indicate that the reliability of the SMI is lowered when overrides are employed. Without the use of overrides, the specific non-acute reliability coefficient for SMI admission reviews between SuperPRO and PRO reviewer 1 is 58%; between SuperPRO and PRO reviewer 2 it is 46%; and between PRO reviewer 1 and PRO reviewer 2 it is 55% (see Table A - 2). These reliability coefficients are all significant (kappa = .62, .52 and .58, $p < .01$, respectively) and begin to approach moderate levels of reliability.

However, it must be emphasized that these reliability coefficients were obtained with the use of extensively trained reviewers and, therefore, are not directly comparable with the reliability coefficients obtained using the AEP criteria on the same set of cases with newly trained AEP auditors. RNs with two days of training in the use of the AEP performed better than SMI reviewers with two years of experience. The inter-rater reliability for the AEP on the same set of 99 cases without the use of overrides is 89.36% (kappa = .78, $p < .01$) in comparison to a specific non-acute agreement rate of 54.55% (kappa = .58, $p < .01$) for the PRO reviewers using the SMI. Thus, the rate of agreement for reviewers using the AEP (with or without overrides) is higher than the specific non-acute agreement produced by reviewers using the SMI.

While these additional reliability reviews indicate that it is possible to improve the reliability of the Standardized Medreview Instrument beyond that obtained with nurses newly trained in its use, the data suggest that the SMI criteria and/or training and training manual may be unclear and/or difficult to comprehend and use in a consistent manner. This is especially true for nurses newly trained in application of the SMI. However, with extensive use and experienced reviewers working in the same organizational setting an improved agreement coefficient is obtained with the SMI. The AEP appears to be superior to the SMI instrument. However, we must temper our earlier conclusions that the SMI is an unreliable set of criteria. The SMI is an instrument that may provide moderately reliable results with nurses extensively trained in its use. However, SMI reliability is below that of the AEP.

Conclusions

While the PRO SMI reliability coefficients for specific non-acute care are higher than those achieved with the use of nurses newly trained in the use of the SMI, they are lower than those obtained with the AEP. Since these reviews were performed under ideal circumstances with extensively trained SMI reviewers from the same organizational setting, we conclude that the reliability of the SMI is lower than the AEP and the SMI should only be used after the reliability of the auditors conducting the reviews is firmly established. Our data suggest that the reliability of SMI reviews, with RNs newly trained in its use, will be inadequate. Since specific non-acute reliability is vital to the development of a valid instrument, and since the AEP is more reliable than the SMI in non-acute reliability, the use of

the SMI should be limited until further refinement of its reliability by developers of the instrument is conducted.*

---

* Blumenfeld, in his critique of the SMI, placed similar cautions on the reliability of the SMI based on developmental methodology. It should also be noted that Blumenfeld (1983) placed similar cautions on the use and development of the AEP (pg. 13-15, and pg. 128-129).

Table A - 1

SMI Reliability Trials with Indiana PRO, SysteMetrics (SuperPRO) and BCBSM Reviewers

Specific Non-Acute Reliability Coefficient (with use of overrides)
Admission Reviews (N=98)

	PRO Reviewer 1	PRO Reviewer 2	SysteMetrics Reviewer	BCBSM Reviewers
PRO Reviewer 1		43.18% Kappa = .43*	44.68% Kappa = .42*	32.56% Kappa = .32*
PRO Reviewer 2			35.90% Kappa = .36*	27.27% Kappa = .27*
SysteMetrics Reviewer				27.03% Kappa = .26*
BCBSM Reviewers				10.34% Kappa = .03

* $p < .01$.

Table A - 2

SMI Reliability Trials with Indiana PRO, SysteMetrics (SuperPro) and BCBSM Reviewers

Specific Non-Acute Reliability Coefficients (without the use of overrides)# Admission Reviews (N=98)

	PRO Reviewer 1	PRO Reviewer 2	SysteMetrics Reviewer	BCBSM Reviewers
PRO Reviewer 1 (1)		54.55% Kappa = .58*	58.06% Kappa = .62*	37.84% Kappa = .38*
PRO Reviewer 2 (2)			46.15% Kappa = .52*	28.13% Kappa = .27*
SysteMetrics Reviewer (3)				31.03% Kappa = .33*
BCBSM Reviewers (4)				13.64% Kappa = .11

Disregarding changes due to the override option. All evaluations are based on application of the objective criteria regardless of the use of the override.

(1) Override use = 12.25%.

(2) Override use = 8.16%.

(3) Override use = 13.27%.

(4) Override use = 17.69%.

* $p < .01$.

APPENDICES A, B and C (BOOKS II AND III) SUPPORTING TABLES

AVAILABLE UPON REQUEST (2 Volumes)

Tue May 24 09:45:26 EDT 2011

b2951757 Last Updated: 05-24-2011 Created: 05-11-2011 Revisions: 5

LANG	engEnglish	COPIES	3	BIB LEVEL	mBOOKS
SKIP	0	CAT DATE	- -	BCODE3	----
LOCATION	multi	MAT TYPE	aPRINT	COUNTRY	miuMichigan

MARC Leader #####cam 22#####Ia 4500
LOCATIONS cmsbk,cms

o	001			32026949
y	008			950221s1987 miua b o000 0 eng d
m	035			(OCoLC)32026949
y	040			EYM|cEYM|dOCLCQ|dHEC
y	043			n-us---
y	049			HECA
c	090			RA394|b.S77 1987
y	090			RA395|b.S77 1987
a	100	1		Strumwasser, Ira.
t	245	1	0	Estimates of non-acute hospitalization :|ba comparative analysis of the appropriateness evaluation protocol and the standardized medreview instrument : final report /|csubmitted by Ira Strumwasser, Nitin V. Paranjpe, and the research staff of the Michigan Health Care Education and Research Foundation, Inc. ... to the Health Care Financing Administration, Office of Research and Development.
p	260			[Detroit, Mich.?] :|bMichigan Health Care Education and Research Foundation,|c[1987]
r	300			3 v. :|bill. ;|c28 cm.
n	500			Supported by HCFA cooperative grant no. 18-C-98582/5-01 & 02.
n	500			"September 1, 1987."
n	504			Includes bibliographical references.
n	505	0		Book I -- Book II. Appendix A (Reliability) and appendix B (Validity) -- Book III. Appendix C (Estimates of non-essential acute care hospitalization)
n	513			Final report;|bJuly 1984-July 1987.
n	530			Available at cost as a print-on-demand technical report ;|bNational Technical Information Service;|dPB89-127237 (Book I)
n	530			Available at cost as a print-on-demand technical report ;|bNational Technical Information Service;|dPB89-127245 (Book II)
n	530			Available at cost as a print-on-demand technical report ;|bNational Technical Information Service;|dPB89-127252 (Book III)
d	650		0	Medical care|xEvaluation|zMichigan.
d	650		0	Hospital utilization|xLength of stay|zMichigan.
d	650		0	Hospitals|xProspective payment|zMichigan.
d	650		0	Diagnosis related groups|zMichigan.
d	650		2	Diagnostic Tests, Routine.
d	650		2	Patient Admission.
b	700	1		Paranjpe, Nitin V.
b	710	1		United States.|bHealth Care Financing Administration.
b	710	2		Michigan Health Care Education and Research Foundation, Inc.

ESTIMATES OF NON-ACUTE HOSPITALIZATION: A COMPARATIVE ANALYSIS OF THE APPROPRIATENESS EVALUATION PROTOCOL AND THE STANDARDIZED MEDREVIEW INSTRUMENT

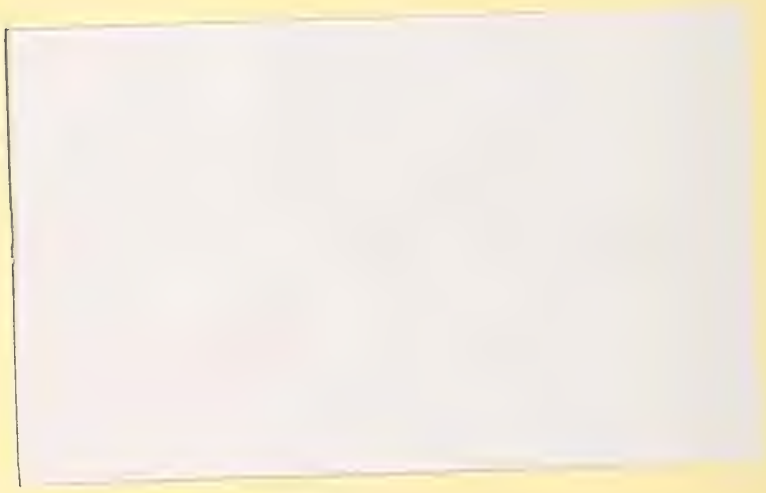
FINAL REPORT - BOOK II OF III

HCFA Grant No. 18-C-98582/5-01 & 02

APPENDIX A (RELIABILITY)

AND

APPENDIX B (VALIDITY)



ESTIMATES OF NON-ACUTE HOSPITALIZATION: A COMPARATIVE ANALYSIS OF THE APPROPRIATENESS EVALUATION PROTOCOL AND THE STANDARDIZED MEDREVIEW INSTRUMENT

FINAL REPORT - BOOK II OF III

HCFA Grant No. 18-C-98582/5-01 & 02

APPENDIX A (RELIABILITY)

AND

APPENDIX B (VALIDITY)

ESTIMATES OF NON-ACUTE HOSPITALIZATION:

A COMPARATIVE ANALYSIS OF THE APPROPRIATENESS EVALUATION PROTOCOL AND THE STANDARDIZED MEDREVIEW INSTRUMENT

FINAL REPORT

submitted by

Ira Strumwasser, Ph.D.
Director of Research
Principal Investigator

Nitin V. Paranjpe, Ph.D.
Senior Research Analyst
Project Director

and

The Research Staff of the

Michigan Health Care Education and Research Foundation, Inc.
(A supporting, non-profit, 501(c)(3), health services research subsidiary of
Health Service Company
a subsidiary of
Blue Cross and Blue Shield of Michigan)

to the

HEALTH CARE FINANCING ADMINISTRATION
OFFICE OF RESEARCH AND DEVELOPMENT

September 1, 1987

HCFA Cooperative Grant No.: 18-C-98582/5-01 & 02

HCFA Project Officer: James Beebe
Office of Research and Development

APPENDIX A

Appendix to Chapter III

Reliability Tables

Admissions

Inter-Rater Reliability

TABLE III - 9

ADMISSIONS

INTER-RATER RELIABILITY

Reliability Coefficients

Protocol	Overall	Specific Non-Acute	Specific Acute	Kappa
AEP	78.72% (74/94)	74.68% (59/79)	42.86% (15/35)	0.47*
SMI	73.47% (72/98)	10.34% (3/29)	72.63% (69/95)	0.03

* Kappa-statistic significant, $p < .01$.

TABLE III - 10

ADMISSIONS

AEP INTER-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review	ACUTE	15	4	19
	NON-ACUTE	16	59	75
	TOTAL	31	63	94

TABLE III - 11

ADMISSIONS

SMI INTER-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review	ACUTE	69	15	84
	NON-ACUTE	11	3	14
	TOTAL	80	18	98

Inter-Rater Reliability Medical Admissions

TABLE III - 12

MEDICAL ADMISSIONS

INTER-RATER RELIABILITY

Reliability Coefficients

Protocol	Overall	Specific Non-Acute	Specific Acute	Kappa
AEP	79.07%	74.65%	45.45%	0.49*
	(68/86)	(53/71)	(15/33)	
SMI	70.79%	10.34%	69.77%	0.01
	(63/89)	(3/29)	(60/86)	

* Kappa-statistic significant, $p < .01$.

TABLE III - 13

MEDICAL ADMISSIONS

AEP INTER-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review	ACUTE	15	4	19
	NON-ACUTE	14	53	67
	TOTAL	29	57	86

TABLE III - 14

MEDICAL ADMISSIONS

SMI INTER-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review	ACUTE	60	15	75
	NON-ACUTE	11	3	14
	TOTAL	71	18	89

Surgical Admissions

Inter-Rater Reliability

TABLE III - 15

SURGICAL ADMISSIONS

INTER-RATER RELIABILITY

Reliability Coefficients

Protocol	Overall	Specific Non-Acute	Specific Acute	Kappa
SAEP*	75.00% (18/24)	60.00% (9/15)	60.00% (9/15)	0.53**
SMI	100.00% (9/9)	- (0/0)	100.00% (9/9)	-

* All 24 elective surgery cases from the validation sample (N = 173) were used in the SAEP reliability trials to increase the original reliability sample from N = 9 to N = 24.

** Kappa-statistic significant, $p < .01$.

TABLE III - 16

SAEP SURGICAL ADMISSIONS

INTER-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review	ACUTE	9	0	9
	NON-ACUTE	6	9	15
	TOTAL	15	9	24

TABLE III - 17

SMI SURGICAL ADMISSIONS

INTER-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review	ACUTE	9	0	9
	NON-ACUTE	0	0	0
	TOTAL	9	0	9

Appropriateness Evaluation Protocol

Inter-Rater Admission Reviews

TABLE III - 18

ADMISSIONS

AEP INTER-RATER RELIABILITY

Reliability Coefficients

Rater	Overall	Specific Non-Acute	Specific Acute	Kappa
A	80.77%	76.19%	50.00%	0.54*
	(21/26)	(16/21)	(5/10)	
B	84.00%	81.82%	42.86%	0.50*
	(21/25)	(18/22)	(3/7)	
C	62.50%	55.00%	30.77%	0.25
	(15/24)	(11/20)	(4/13)	
D	75.00%	70.00%	40.00%	0.44*
	(18/24)	(14/20)	(4/10)	

* Kappa-statistic significant, $p < .01$.

TABLE III - 19

ADMISSIONS

AEP INTER-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review (Rater A)	ACUTE	5	1	6
	NON-ACUTE	4	16	20
	TOTAL	9	17	26

TABLE III - 20

ADMISSIONS

AEP INTER-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review (Rater B)	ACUTE	3	2	5
	NON-ACUTE	2	18	20
	TOTAL	5	20	25

TABLE III - 21

ADMISSIONS

AEP INTER-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review (Rater C)	ACUTE	4	1	5
	NON-ACUTE	8	11	19
	TOTAL	12	12	24

TABLE III - 22

ADMISSIONS

AEP INTER-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review (Rater D)	ACUTE	4	0	4
	NON-ACUTE	6	14	20
	TOTAL	10	14	24

Standardized Medreview Instrument

Inter-Rater Admission Reviews

TABLE III - 23

ADMISSIONS

SMI INTER-RATER RELIABILITY

Reliability Coefficients

Rater	Overall	Specific Non-Acute	Specific Acute	Kappa
A	66.67%	25.00%	62.50%	0.27
	(6/9)	(1/4)	(5/8)	
B	71.43%	11.11%	70.37%	0.03
	(40/56)	(2/18)	(38/54)	
C	83.33%	00.00%	83.33%	
	(15/18)	(0/3)	(15/18)	
D	73.33%	00.00%	73.33%	-0.11
	(11/15)	(0/4)	(11/15)	

TABLE III - 24

ADMISSIONS

SMI INTER-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review (Rater A)	ACUTE	5	0	5
	NON-ACUTE	3	1	4
	TOTAL	8	1	9

TABLE III - 25

ADMISSIONS

SMI INTER-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review (Rater B)	ACUTE	38	9	47
	NON-ACUTE	7	2	9
	TOTAL	45	11	56

TABLE III - 26

ADMISSIONS

SMI INTER-RATER RELIABILITY

(Pattern of Agreement)

Re-Review

		ACUTE	NON-ACUTE	TOTAL
Initial Review (Rater C)	ACUTE	15	3	18
	NON-ACUTE	0	0	0
	TOTAL	15	3	18

TABLE III - 27

ADMISSIONS

SMI INTER-RATER RELIABILITY

(Pattern of Agreement)

Re-Review

		ACUTE	NON-ACUTE	TOTAL
Initial Review (Rater D)	ACUTE	11	3	14
	NON-ACUTE	1	0	1
	TOTAL	12	3	15

Days Of Care

Inter-Rater Reliability

TABLE III - 28

DAYS OF CARE

INTER-RATER RELIABILITY

Reliability Coefficients

Protocol	Overall	Specific Non-Acute	Specific Acute	Kappa
AEP	83.84%	80.40%	52.03%	0.58*
	(306/365)	(242/301)	(64/123)	
SMI	66.75%	34.54%	59.68%	0.27*
	(255/382)	(67/194)	(188/315)	

* Kappa-statistic significant, $p < .01$.

TABLE III - 29

DAYS OF CARE

AEP INTER-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review	ACUTE	64	30	94
	NON-ACUTE	29	242	271
	TOTAL	93	272	365

TABLE III - 30

DAYS OF CARE

SMI INTER-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review	ACUTE	67	45	112
	NON-ACUTE	82	188	270
	TOTAL	149	233	382

Appropriateness Evaluation Protocol:

Days Of Care Review

TABLE III - 31

DAYS OF CARE

AEP INTER-RATER RELIABILITY

Reliability Coefficients

Rater	Overall	Specific Non-Acute	Specific Acute	Kappa
A	80.43%	76.00%	48.57%	0.52*
	(74/92)	(57/75)	(17/35)	
B	83.70%	79.45%	55.88%	0.60*
	(77/92)	(58/73)	(19/34)	
C	80.90%	78.48%	37.04%	0.43*
	(72/89)	(62/79)	10/27)	
D	86.46%	83.33%	58.06%	0.64*
	(83/96)	(65/78)	(18/31)	

* Kappa-statistic significant, $p < .01$.

TABLE III - 32

DAYS OF CARE

AEP INTER-RATER RELIABILITY

(Pattern of Agreement)

Re-Review

		ACUTE	NON-ACUTE	TOTAL
Initial Review (Rater A)	ACUTE	17	13	30
	NON-ACUTE	5	57	62
	TOTAL	22	70	92

TABLE III - 33

DAYS OF CARE

AEP INTER-RATER RELIABILITY

(Pattern of Agreement)

Re-Review

		ACUTE	NON-ACUTE	TOTAL
Initial Review (Rater B)	ACUTE	19	5	24
	NON-ACUTE	10	58	68
	TOTAL	29	63	92

TABLE III - 34

DAYS OF CARE

AEP INTER-RATER RELIABILITY

(Pattern of Agreement)

Re-Review

		ACUTE	NON-ACUTE	TOTAL
Initial Review (Rater C)	ACUTE	10	5	15
	NON-ACUTE	12	62	74
	TOTAL	22	67	89

TABLE III - 35

DAYS OF CARE

AEP INTER-RATER RELIABILITY

(Pattern of Agreement)

Re-Review

		ACUTE	NON-ACUTE	TOTAL
Initial Review (Rater D)	ACUTE	18	7	25
	NON-ACUTE	6	65	71
	TOTAL	24	72	96

Standardized Medreview Instrument:

Days Of Care Review

TABLE III - 36

DAYS OF CARE

SMI INTER-RATER RELIABILITY

Reliability Coefficients

Rater	Overall	Specific Non-Acute	Specific Acute	Kappa
A	89.65%	72.73%	85.71%	0.77*
	(26/29)	(8/11)	(18/21)	
B	62.56%	31.09%	54.95%	0.21*
	(137/219)	(37/119)	(100/182)	
C	61.64%	31.71%	53.33%	0.18
	(45/73)	(13/41)	(32/60)	
D	71.67%	37.04%	66.00%	0.34*
	(43/60)	(10/27)	(33/50)	

* Kappa-statistic significant, $p < .01$.

TABLE III - 37

DAYS OF CARE

SMI INTER-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review (Rater A)	ACUTE	8	2	10
	NON-ACUTE	1	18	19
	TOTAL	9	20	29

TABLE III - 38

DAYS OF CARE

SMI INTER-RATER RELIABILITY

(Pattern of Agreement)

Re-Review

		ACUTE	NON-ACUTE	TOTAL
Initial Review (Rater B)	ACUTE	37	23	60
	NON-ACUTE	59	100	159
	TOTAL	96	123	219

TABLE III - 39

DAYS OF CARE

SMI INTER-RATER RELIABILITY

(Pattern of Agreement)

Re-Review

		ACUTE	NON-ACUTE	TOTAL
Initial Review (Rater C)	ACUTE	13	14	27
	NON-ACUTE	14	32	46
	TOTAL	27	46	73

TABLE III - 40

DAYS OF CARE

SMI INTER-RATER RELIABILITY

(Pattern of Agreement)

Re-Review

		ACUTE	NON-ACUTE	TOTAL
Initial Review (Rater D)	ACUTE	10	7	17
	NON-ACUTE	10	33	43
	TOTAL	20	40	60

Admissions:

Intra-Rater Reliability

TABLE III - 41

ADMISSIONS

INTRA-RATER RELIABILITY

Reliability Coefficients

Protocol	Overall	Specific Non-Acute	Specific Acute	Kappa
AEP	86.02%	83.12%	55.17%	0.62*
	(80/93)	(64/77)	(16/29)	
SMI	78.72%	28.57%	76.74%	0.32*
	(74/94)	(8/28)	(66/86)	

* Kappa-statistic significant, $p < .01$.

TABLE III - 42

ADMISSIONS

AEP INTRA-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review	ACUTE	16	3	19
	NON-ACUTE	10	64	74
	TOTAL	26	67	93

TABLE III - 43

ADMISSIONS

SMI INTRA-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review	ACUTE	66	14	80
	NON-ACUTE	6	8	14
	TOTAL	72	22	94

Appropriateness Evaluation Protocol:

Intra-Rater Admission Reviews

TABLE III - 44

ADMISSIONS

AEP INTRA-RATER RELIABILITY

Reliability Coefficients

Rater	Overall	Specific Non-Acute	Specific Acute	Kappa
A	91.30% (21/23)	88.89% (16/18)	71.43% (5/7)	0.78*
B	80.00% (20/25)	75.00% (15/20)	50.00% (5/10)	0.55*
C	87.50% (21/24)	85.00% (17/20)	57.14% (4/7)	0.65*
D	71.43% (15/21)	66.67% (12/18)	33.33% (3/9)	0.33

* Kappa-statistic significant, $p < .01$.

TABLE III - 45

ADMISSIONS

AEP INTRA-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review (Rater A)	ACUTE	5	1	6
	NON-ACUTE	1	16	17
	TOTAL	6	17	23

TABLE III - 46

ADMISSIONS

AEP INTRA-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review (Rater B)	ACUTE	5	0	5
	NON-ACUTE	5	15	20
	TOTAL	10	15	25

TABLE III - 47

ADMISSIONS

AEP INTRA-RATER RELIABILITY

(Pattern of Agreement)

Re-Review

		ACUTE	NON-ACUTE	TOTAL
Initial Review (Rater C)	ACUTE	4	1	5
	NON-ACUTE	2	17	19
	TOTAL	6	18	24

TABLE III - 48

ADMISSIONS

AEP INTRA-RATER RELIABILITY

(Pattern of Agreement)

Re-Review

		ACUTE	NON-ACUTE	TOTAL
Initial Review (Rater D)	ACUTE	3	1	4
	NON-ACUTE	5	12	17
	TOTAL	8	13	21

Standardized Medreview Instrument:

Intra-Rater Admission Reviews

TABLE III - 49

ADMISSIONS

SMI INTRA-RATER RELIABILITY

PAIRWISE COMPARISONS

Reliability Coefficients

Rater	Overall	Specific Non-Acute	Specific Acute	Kappa
A	77.78%	50.00%	71.43%	0.53
	(7/9)	(2/4)	(5/7)	
B	76.36%	27.78%	74.00%	0.29*
	(42/55)	(5/18)	(37/50)	
C	94.12%	00.00%	94.12%	-
	(16/17)	(0/1)	(16/17)	
D	80.00%	25.00%	78.57 %	0.33
	(12/15)	(1/4)	(11/14)	

* Kappa-statistic significant, $p < .01$.

TABLE III - 50

ADMISSIONS

SMI INTRA-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review (Rater A)	ACUTE	5	0	5
	NON-ACUTE	2	2	4
	TOTAL	7	2	9

TABLE III - 51

ADMISSIONS

SMI INTRA-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review (Rater B)	ACUTE	37	9	46
	NON-ACUTE	4	5	9
	TOTAL	41	14	55

TABLE III - 52

ADMISSIONS

SMI INTRA-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review (Rater C)	ACUTE	16	1	17
	NON-ACUTE	0	0	0
	TOTAL	16	1	17

TABLE III - 53

ADMISSIONS

SMI INTRA-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review (Rater D)	ACUTE	11	3	14
	NON-ACUTE	0	1	1
	TOTAL	11	4	15

Days of Care

Intra-Rater Reliability

TABLE III - 54

DAYS OF CARE

INTRA-RATER RELIABILITY

Reliability Coefficients

Protocol	Overall	Specific Non-Acute	Specific Acute	Kappa
AEP	88.25% (323/366)	85.42% (252/295)	62.28% (71/114)	0.69*
SMI	74.03% (285/385)	46.81% (88/188)	66.33% (197/297)	0.45*

* Kappa-statistic significant, $p < .01$.

TABLE III - 55

DAYS OF CARE

AEP INTRA-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review	ACUTE	71	23	94
	NON-ACUTE	20	252	272
	TOTAL	91	275	366

TABLE III - 56

DAYS OF CARE

SMI INTRA-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review	ACUTE	88	24	112
	NON-ACUTE	76	197	273
	TOTAL	164	221	385

Appropriateness Evaluation Protocol

Intra-Rater DOS Reliability

TABLE III - 57

DAYS OF CARE

AEP INTRA-RATER RELIABILITY

Reliability Coefficients

Rater	Overall	Specific Non-Acute	Specific Acute	Kappa
A	91.67%	89.39%	72.00%	0.78*
	(77/84)	(59/66)	(18/25)	
B	81.52%	77.33%	50.00%	0.54*
	(75/92)	(58/75)	(17/34)	
C	98.86%	98.63%	93.75%	0.96*
	(87/88)	(72/73)	(15/16)	
D	79.52%	74.24%	50.00 %	0.52*
	(66/83)	(49/66)	(17/34)	

* Kappa-statistic significant, $p < .01$.

TABLE III - 58

DAYS OF CARE

AEP INTRA-RATER RELIABILITY

(Pattern of Agreement)

Re-Review

		ACUTE	NON-ACUTE	TOTAL
Initial Review (Rater A)	ACUTE	18	7	25
	NON-ACUTE	0	59	59
	TOTAL	18	66	84

TABLE III - 59

DAYS OF CARE

AEP INTRA-RATER RELIABILITY

(Pattern of Agreement)

Re-Review

		ACUTE	NON-ACUTE	TOTAL
Initial Review (Rater B)	ACUTE	17	7	24
	NON-ACUTE	10	58	68
	TOTAL	27	65	92

TABLE III - 60

DAYS OF CARE

AEP INTRA-RATER RELIABILITY

(Pattern of Agreement)

Re-Review

		ACUTE	NON-ACUTE	TOTAL
Initial Review (Rater C)	ACUTE	15	0	15
	NON-ACUTE	1	72	73
	TOTAL	16	72	88

TABLE III - 61

DAYS OF CARE

AEP INTRA-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review (Rater D)	ACUTE	17	8	25
	NON-ACUTE	9	49	58
	TOTAL	26	57	83

Standardized Medreview Instrument:

Intra-Rater DOS Reliability

TABLE III - 62

DAYS OF CARE

SMI INTRA-RATER RELIABILITY

Reliability Coefficients

Rater	Overall	Specific Non-Acute	Specific Acute	Kappa
A	86.21%	63.64%	81.82%	0.68*
	(25/29)	(7/11)	(18/22)	
B	68.95%	41.88%	60.00%	0.37*
	(151/219)	(49/117)	(102/170)	
C	76.12%	57.89%	64.44%	0.52*
	(51/67)	(22/38)	(29/45)	
D	80.33%	52.00%	75.00 %	0.54*
	(49/61)	(13/25)	(36/48)	

* Kappa-statistic significant, $p < .01$.

TABLE III - 63

DAYS OF CARE

SMI INTRA-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review (Rater A)	ACUTE	7	3	10
	NON-ACUTE	1	18	19
	TOTAL	8	21	29

TABLE III - 64

DAYS OF CARE

SMI INTRA-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review (Rater B)	ACUTE	49	11	60
	NON-ACUTE	57	102	159
	TOTAL	106	113	219

TABLE III - 65

DAYS OF CARE

SMI INTRA-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review (Rater C)	ACUTE	22	6	28
	NON-ACUTE	10	29	39
	TOTAL	32	35	67

TABLE III - 66

DAYS OF CARE

SMI INTRA-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review (Rater D)	ACUTE	13	5	18
	NON-ACUTE	7	36	43
	TOTAL	20	41	61

Medical Days of Care

Inter-Rater Reliability

TABLE III - 67

MEDICAL DAYS OF CARE

INTER-RATER RELIABILITY

Reliability Coefficients

Protocol	Overall	Specific Non-Acute	Specific Acute	Kappa
AEP	84.07% (285/339)	81.25% (234/288)	48.57% (51/105)	0.55*
SMI	65.83% (235/357)	35.45% (67/189)	57.93% (168/290)	0.26*

* Kappa-statistic significant, $p < .01$.

TABLE III - 68

MEDICAL DAYS OF CARE

AEP INTER-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review	ACUTE	51	30	81
	NON-ACUTE	24	234	258
	TOTAL	75	264	339

TABLE III - 69

MEDICAL DAYS OF CARE

SMI INTER-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review	ACUTE	168	77	245
	NON-ACUTE	45	67	112
	TOTAL	213	144	357

Days Within Surgical Admissions

Inter-Rater Reliability

TABLE III - 70

DAYS WITHIN SURGICAL ADMISSIONS*

INTER-RATER RELIABILITY

Protocol	Reliability Coefficients Overall	Specific Non-Acute	Specific Acute	Kappa
AEP	69.07%	47.37%	57.14%	0.38**
	(67/97)	(27/57)	(40/70)	

* N based on 24 admissions.

** Kappa-statistic significant, $p < .01$.

TABLE III - 71

DAYS WITHIN SURGICAL ADMISSIONS

AEP INTER-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review	ACUTE	40	10	50
	NON-ACUTE	20	27	47
	TOTAL	60	37	97

Medical Admissions

Intra-Rater Reliability

TABLE III - 72

MEDICAL ADMISSIONS

INTRA-RATER RELIABILITY

Protocol	Reliability Coefficients: Overall	Specific Non-Acute	Specific Acute	Kappa
AEP	86.90%	83.82%	59.26%	0.66*
	(73/84)	(57/68)	(16/27)	
SMI	76.47%	28.57%	74.03%	0.30*
	(65/85)	(8/28)	(57/77)	

* Kappa-statistic significant, $p < .01$.

TABLE III - 73

AEP MEDICAL ADMISSIONS

INTRA-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review	ACUTE	16	3	19
	NON-ACUTE	8	57	65
	TOTAL	24	60	84

TABLE III - 74

SMI MEDICAL ADMISSIONS

INTRA-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review	ACUTE	57	14	71
	NON-ACUTE	6	8	14
	TOTAL	63	22	85

Medical Days of Care

Intra-Rater Reliability

TABLE III - 75

MEDICAL DAYS OF CARE

INTRA-RATER RELIABILITY

Reliability Coefficients

Protocol	Overall	Specific Non-Acute	Specific Acute	Kappa
AEP	88.24% (300/340)	85.82% (242/282)	59.18% (58/98)	0.67*
SMI	72.78% (262/360)	47.31% (88/186)	63.97% (174/272)	0.43*

* Kappa-statistic significant, $p < .01$.

TABLE III - 76

AEP MEDICAL DAYS OF CARE

INTRA-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review	ACUTE	58	23	81
	NON-ACUTE	17	242	259
	TOTAL	75	265	340

TABLE III - 77

SMI MEDICAL DAYS OF CARE

INTRA-RATER RELIABILITY

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review	ACUTE	174	74	248
	NON-ACUTE	24	88	112
	TOTAL	198	162	360

Inter-Rater Reliability

Without the Use of Overrides

TABLE III - 78

INTER-RATER RELIABILITY ON ADMISSIONS

WITHOUT THE USE OF OVERRIDES

Reliability Coefficients

Protocol	Overall	Specific Non-Acute	Specific Acute	Kappa
AEP	91.67%	89.36%	72.22%	0.78*
	(55/60)	(42/47)	(13/18)	
SMI	78.41%	13.64%	77.65%	0.11
	(69/88)	(3/22)	(66/85)	

* Kappa-statistic significant, $p < .01$.

TABLE III - 79

AEP INTER-RATER RELIABILITY ON ADMISSIONS

WITHOUT THE USE OF OVERRIDES

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review	ACUTE	13	1	14
	NON-ACUTE	4	42	46
	TOTAL	17	43	60

TABLE III - 80

SMI INTER-RATER RELIABILITY ON ADMISSIONS

WITHOUT THE USE OF OVERRIDES

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review	ACUTE	66	10	76
	NON-ACUTE	9	3	12
	TOTAL	75	13	88

TABLE III - 81

INTER-RATER RELIABILITY FOR DAY REVIEWS

WITHOUT THE USE OF OVERRIDES

Reliability Coefficients

Protocol	Overall	Specific Non-Acute	Specific Acute	Kappa
AEP	89.81% (238/265)	86.63% (175/202)	70.00% (63/90)	0.75*
SMI	67.02% (250/373)	34.92% (66/189)	59.93% (184/307)	0.28*

* Kappa-statistic significant, $p < .01$.

TABLE III - 82

AEP INTER-RATER RELIABILITY FOR DAY REVIEWS

WITHOUT THE USE OF OVERRIDES

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review	ACUTE	63	21	84
	NON-ACUTE	6	175	181
	TOTAL	69	196	265

TABLE III - 83

SMI INTER-RATER RELIABILITY FOR DAY REVIEWS

WITHOUT THE USE OF OVERRIDES

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review	ACUTE	66	43	109
	NON-ACUTE	80	184	264
	TOTAL	146	227	373

Intra-Rater Reliability

Without the Use of Overrides

TABLE III - 84

INTRA-RATER RELIABILITY FOR ADMISSIONS

WITHOUT THE USE OF OVERRIDES

	Reliability Coefficients			
Protocol	Overall	Specific Non-Acute	Specific Acute	Kappa
AEP	92.42% (61/66)	90.20% (46/51)	75.00% (15/20)	0.81*
SMI	82.35% (70/85)	34.78% (8/23)	80.52% (62/77)	0.41*

* Kappa-statistic significant, $p < .01$.

TABLE III - 85

AEP INTRA-RATER RELIABILITY FOR ADMISSIONS

WITHOUT THE USE OF OVERRIDES

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review	ACUTE	15	1	16
	NON-ACUTE	4	46	50
	TOTAL	19	47	66

TABLE III - 86

SMI INTRA-RATER RELIABILITY FOR ADMISSIONS

WITHOUT THE USE OF OVERRIDES

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review	ACUTE	62	10	72
	NON-ACUTE	5	8	13
	TOTAL	67	18	85

TABLE III - 87

INTRA-RATER RELIABILITY FOR DAY REVIEWS

WITHOUT THE USE OF OVERRIDES

Reliability Coefficients

Protocol	Overall	Specific Non-Acute	Specific Acute	Kappa
AEP	93.90% (231/246)	91.57% (163/178)	81.93% (68/83)	0.86*
SMI	74.08% (283/382)	47.06% (88/187)	66.33% (195/294)	0.45*

* Kappa-statistic significant, $p < .01$.

TABLE III - 88

AEP INTRA-RATER RELIABILITY FOR DAY REVIEWS

WITHOUT THE USE OF OVERRIDES

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review	ACUTE	68	10	78
	NON-ACUTE	5	163	168
	TOTAL	73	173	246

TABLE III - 89

SMI INTRA-RATER RELIABILITY FOR DAY REVIEWS

WITHOUT THE USE OF OVERRIDES

(Pattern of Agreement)

		Re-Review		
		ACUTE	NON-ACUTE	TOTAL
Initial Review	ACUTE	88	23	111
	NON-ACUTE	76	195	271
	TOTAL	164	218	382

APPENDIX B

Appendix to Chapter III

Validity Tables

FFS Physician Validation

TABLE III - 90

MEDICAL ADMISSIONS

AEP/FFS PHYSICIAN VALIDITY

Validity Coefficients

Rule	Overall	Specific Non-Acute	Kappa
1: Lenient	75.67% (112/148)	70.73% (87/123)	0.43*
2: Moderate	48.65% (72/148)	36.13% (43/119)	0.16*
3: Conservative	32.43% (48/148)	14.53% (17/117)	0.07*

* Kappa-statistic significant, $p < .01$.

TABLE III - 91

MEDICAL ADMISSIONS

AEP VALIDITY: LENIENT NON-ACUTE RULE

FFS PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	25	6	31
NON-ACUTE	30	87	117
TOTAL	55	93	148

TABLE III - 92

MEDICAL ADMISSIONS

AEP VALIDITY: MODERATE NON-ACUTE RULE

FFS PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	29	2	31
NON-ACUTE	74	43	117
TOTAL	103	45	148

TABLE III - 93

MEDICAL ADMISSIONS

AEP VALIDITY: CONSERVATIVE NON-ACUTE RULE

FFS PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	31	0	31
NON-ACUTE	100	17	117
TOTAL	131	17	148

TABLE III - 94

MEDICAL DAYS OF CARE

AEP/FFS PHYSICIAN VALIDITY

Validity Coefficients

Rule	Overall	Specific Non-Acute	Kappa
1: Lenient	76.10% (551/724)	71.02% (424/597)	0.43*
2: Moderate	64.64% (468/724)	54.29% (304/560)	0.30*
3: Conservative	45.30% (328/724)	25.70% (137/533)	0.12*

* Kappa-statistic significant, $p < .01$.

TABLE III - 95

MEDICAL DAYS OF CARE

AEP VALIDITY: LENIENT NON-ACUTE RULE

		FFS PHYSICIANS' ASSESSMENT		
		ACUTE	NON-ACUTE	TOTAL
AEP	ACUTE	127	81	208
	NON-ACUTE	92	424	516
	TOTAL	219	505	724

TABLE III - 96

MEDICAL DAYS OF CARE

AEP VALIDITY: MODERATE NON-ACUTE RULE

FFS PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	164	44	208
NON-ACUTE	212	304	516
TOTAL	376	348	724

TABLE III - 97

MEDICAL DAYS OF CARE

AEP VALIDITY: CONSERVATIVE NON-ACUTE RULE

FFS PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	191	17	208
NON-ACUTE	379	137	516
TOTAL	570	154	724

TABLE III - 98

SURGICAL ADMISSIONS

SAEP/FFS PHYSICIAN VALIDITY

Validity Coefficients

Rule	Overall	Specific Non-Acute	Kappa
1: Lenient	62.50%	47.06%	0.26
	(15/24)	(8/17)	
2: Moderate	54.17%	26.67%	0.20
	(13/24)	(4/15)	
3: Conservative	41.67%	6.67%	0.05
	(10/24)	(1/15)	

TABLE III - 99

SURGICAL ADMISSIONS

SAEP VALIDITY: LENIENT NON-ACUTE RULE

		FFS PHYSICIANS' ASSESSMENT		
		ACUTE	NON-ACUTE	TOTAL
SAEP	ACUTE	7	2	9
	NON-ACUTE	7	8	15
	TOTAL	14	10	24

TABLE III - 100

SURGICAL ADMISSIONS

SAEP VALIDITY: MODERATE NON-ACUTE RULE

FFS PHYSICIANS' ASSESSMENT

SAEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	9	0	9
NON-ACUTE	11	4	15
TOTAL	20	4	24

TABLE III - 101

SURGICAL ADMISSIONS

SAEP VALIDITY: CONSERVATIVE NON-ACUTE RULE

FFS PHYSICIANS' ASSESSMENT

SAEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	9	0	9
NON-ACUTE	14	1	15
TOTAL	23	1	24

TABLE III - 102

DAYS WITHIN SURGICAL ADMISSIONS

AEP/FFS PHYSICIAN VALIDITY

Rule	Validity Coefficients: Overall	Specific Non-Acute	Kappa
1: Lenient	60.82% (59/97)	44.12% (30/68)	0.22*
2: Moderate	58.76% (57/97)	24.53% (13/53)	0.16
3: Conservative	54.64% (53/97)	8.33% (4/48)	0.07

* Kappa-statistic significant, $p < .01$.

TABLE III - 103

DAYS WITHIN SURGICAL ADMISSIONS

AEP VALIDITY: LENIENT NON-ACUTE RULE

FFS PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	26	24	50
NON-ACUTE	16	31	47
TOTAL	42	55	97

TABLE III - 104

DAYS WITHIN SURGICAL ADMISSIONS

AEP VALIDITY: MODERATE NON-ACUTE RULE

FFS PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	44	6	50
NON-ACUTE	34	13	47
TOTAL	78	19	97

TABLE III - 105

DAYS WITHIN SURGICAL ADMISSIONS

AEP VALIDITY: CONSERVATIVE NON-ACUTE RULE

FFS PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	49	1	50
NON-ACUTE	43	4	47
TOTAL	92	5	97

HMO Physician Validation

TABLE III - 106

MEDICAL ADMISSIONS

AEP/HMO PHYSICIAN VALIDITY

Validity Coefficients

Rule	Overall	Specific Non-Acute	Kappa
1: Lenient	81.08% (120/148)	77.24% (95/123)	0.52*
2: Moderate	68.92% (102/148)	61.67% (74/120)	0.36*
3: Conservative	35.14% (52/148)	19.33% (23/119)	0.06

* Kappa-statistic significant, $p < .01$.

TABLE III - 107

MEDICAL ADMISSIONS

AEP/HMO VALIDITY

Lenient Non-Acute Rule

HMO PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	25	6	31
NON-ACUTE	22	95	117
TOTAL	47	101	148

TABLE III - 108

MEDICAL ADMISSIONS

AEP/HMO VALIDITY

Moderate Non-Acute Rule

HMO PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	28	3	31
NON-ACUTE	43	74	117
TOTAL	71	77	148

TABLE III - 109

MEDICAL ADMISSIONS

AEP/HMO VALIDITY

Conservative Non-Acute Rule

HMO PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	29	2	31
NON-ACUTE	94	23	117
TOTAL	123	25	148

TABLE III - 110

MEDICAL DAYS OF CARE

AEP/HMO PHYSICIAN VALIDITY

Validity Coefficients

Rule	Overall	Specific Non-Acute	Kappa
1: Lenient	77.15% (564/731)	73.91% (473/640)	0.38*
2: Moderate	72.78% (532/731)	65.87% (384/583)	0.40*
3: Conservative	53.76% (393/731)	38.66% (213/551)	0.20*

* Kappa-statistic significant, $p < .01$.

TABLE III - 111

MEDICAL DAYS OF CARE

AEP/HMO VALIDITY

Lenient Non-Acute Rule

HMO PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	91	110	201
NON-ACUTE	57	473	530
TOTAL	148	583	731

TABLE III - 112

MEDICAL DAYS OF CARE

AEP/HMO VALIDITY

Moderate Non-Acute Rule

HMO PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	148	53	201
NON-ACUTE	146	384	530
TOTAL	294	437	731

TABLE III - 113

MEDICAL DAYS OF CARE

AEP/HMO VALIDITY

Conservative Non-Acute Rule

HMO PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	180	21	201
NON-ACUTE	317	213	530
TOTAL	497	234	731

TABLE III - 114

SURGICAL ADMISSIONS

SAEP/HMO PHYSICIAN VALIDITY

	Validity Coefficients		
Rule	Overall	Specific Non-Acute	Kappa
1: Lenient	62.50% (15/24)	60.87% (14/23)	0.05
2: Moderate	62.50% (15/24)	57.14% (12/21)	0.14
3: Conservative	70.83% (17/24)	56.25% (9/16)	0.44*

* Kappa-statistic significant, $p < .01$.

TABLE III - 115

SURGICAL ADMISSIONS

SAEP/HMO VALIDITY

Lenient Non-Acute Rule

HMO PHYSICIANS' ASSESSMENT

SAEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	1	8	9
NON-ACUTE	1	14	15
TOTAL	2	22	24

TABLE III - 116

SURGICAL ADMISSIONS

SAEP/HMO VALIDITY

Moderate Non-Acute Rule

		HMO PHYSICIANS' ASSESSMENT		
		ACUTE	NON-ACUTE	TOTAL
SAEP	ACUTE	3	6	9
	NON-ACUTE	3	12	15
	TOTAL	6	18	24

TABLE III - 117

SURGICAL ADMISSIONS

SAEP/HMO VALIDITY

Conservative Non-Acute Rule

HMO PHYSICIANS' ASSESSMENT

SAEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	8	1	9
NON-ACUTE	6	9	15
TOTAL	14	10	24

TABLE III - 118

DAYS WITHIN SURGICAL ADMISSIONS

AEP/HMO PHYSICIAN VALIDITY

	Validity Coefficients		
Rule	Overall	Specific Non-Acute	Kappa
1: Lenient	70.10% (68/97)	59.15% (42/71)	0.41*
2: Moderate	68.04% (66/97)	48.33% (29/60)	0.36*
3: Conservative	62.89% (61/97)	32.08% (17/53)	0.25*

* Kappa-statistic significant, $p < .01$.

TABLE III - 119

DAYS WITHIN SURGICAL ADMISSIONS

AEP/HMO VALIDITY

Lenient Non-Acute Rule

HMO PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	26	24	50
NON-ACUTE	5	42	47
TOTAL	31	66	97

TABLE III - 120

DAYS WITHIN SURGICAL ADMISSIONS

AEP/HMO VALIDITY

Moderate Non-Acute Rule

HMO PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	37	13	50
NON-ACUTE	18	29	47
TOTAL	55	42	97

TABLE III - 121

DAYS WITHIN SURGICAL ADMISSIONS

AEP/HMO VALIDITY

Conservative Non-Acute Rule

HMO PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	44	6	50
NON-ACUTE	30	17	47
TOTAL	74	23	97

FFS/HMO Physician Validation Summary

TABLE III - 122

HMO AND FFS VALIDITY

NON-ACUTE CARE

AEP

MEDICAL	VALIDITY RULE	HMO PERCENT	FFS PERCENT
ADMISSIONS	LIBERAL	77.24	70.73
ADMISSIONS	MODERATE	61.67	36.13
ADMISSIONS	CONSERVATIVE	19.33	14.53
DAYS OF CARE	LIBERAL	73.91	71.02
DAYS OF CARE	MODERATE	65.87	54.29
DAYS OF CARE	CONSERVATIVE	38.66	25.70

TABLE III - 123

HMO AND FFS VALIDITY

NON-ACUTE CARE

AEP

SURGICAL	VALIDITY RULE	HMO PERCENT	FFS PERCENT
ADMISSIONS	LIBERAL	60.87	47.06
ADMISSIONS	MODERATE	57.14	26.67
ADMISSIONS	CONSERVATIVE	56.25	6.67
DAYS OF CARE	LIBERAL	59.15	44.12
DAYS OF CARE	MODERATE	48.33	24.53
DAYS OF CARE	CONSERVATIVE	32.08	8.33

FFS Physician Validation:

AEP No Overrides

TABLE III - 124

MEDICAL ADMISSIONS NO OVERRIDE

AEP/FFS PHYSICIAN VALIDITY

Validity Coefficients

Rule	Overall	Specific Non-Acute	Kappa
1: Lenient	79.34% (96/121)	74.23% (72/97)	0.52*
2: Moderate	54.55% (66/121)	41.49% (39/124)	0.22*
3: Conservative	36.37% (44/121)	16.30% (15/92)	0.08*

* Kappa-statistic significant, $p < .01$.

TABLE III - 125

MEDICAL ADMISSIONS NO OVERRIDE

AEP VALIDITY

Lenient Non-Acute Rule

FFS PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	24	5	29
NON-ACUTE	20	72	92
TOTAL	44	77	121

TABLE III - 126

MEDICAL ADMISSIONS NO OVERRIDE

AEP VALIDITY

Moderate Non-Acute Rule

FFS PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	27	2	29
NON-ACUTE	53	39	92
TOTAL	80	41	121

TABLE III - 127

MEDICAL ADMISSIONS NO OVERRIDE

AEP VALIDITY

Conservative Non-Acute Rule

FFS PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	29	0	29
NON-ACUTE	77	15	92
TOTAL	106	15	121

TABLE III - 128

MEDICAL DAYS OF CARE NO OVERRIDE

AEP/FFS PHYSICIAN VALIDITY

Validity Coefficients

Rule	Overall	Specific Non-Acute	Kappa
1: Lenient	76.94% (427/555)	70.09% (300/528)	0.49*
2: Moderate	69.73% (387/555)	57.03% (223/391)	0.40*
3: Conservative	52.61% (292/555)	28.14% (103/366)	0.18*

* Kappa-statistic significant, $p < .01$.

TABLE III - 129

MEDICAL DAYS OF CARE NO OVERRIDE

AEP VALIDITY

Lenient Non-Acute Rule

FFS PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	127	77	204
NON-ACUTE	51	300	351
TOTAL	178	377	555

TABLE III - 130

MEDICAL DAYS OF CARE NO OVERRIDE

AEP VALIDITY

Moderate Non-Acute Rule

FFS PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	164	40	204
NON-ACUTE	128	223	351
TOTAL	292	263	555

TABLE III - 131

MEDICAL DAYS OF CARE NO OVERRIDE

AEP VALIDITY

Conservative Non-Acute Rule

FFS PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	189	15	204
NON-ACUTE	248	103	351
TOTAL	437	118	555

TABLE III - 132

DAYS WITHIN SURGICAL ADMISSIONS NO OVERRIDE

AEP/FFS PHYSICIAN VALIDITY

	Validity Coefficients		
Rule	Overall	Specific Non-Acute	Kappa
1: Lenient	64.79% (46/71)	44.44% (20/45)	0.34*
2: Moderate	76.05% (54/71)	41.38% (12/29)	0.42*
3: Conservative	71.83% (51/71)	16.67% (4/24)	0.19*

* Kappa-statistic significant, $p < .01$.

TABLE III - 133

DAYS WITHIN SURGICAL ADMISSIONS NO OVERRIDE

AEP VALIDITY

Lenient Non-Acute Rule

FFS PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	26	22	48
NON-ACUTE	3	20	23
TOTAL	29	42	71

TABLE III - 134

DAYS WITHIN SURGICAL ADMISSIONS NO OVERRIDE

AEP VALIDITY

Moderate Non-Acute Rule

FFS PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	42	6	48
NON-ACUTE	11	12	23
TOTAL	53	18	71

TABLE III - 135

DAYS WITHIN SURGICAL ADMISSIONS NO OVERRIDE

AEP VALIDITY

Conservative Non-Acute Rule

FFS PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	47	1	48
NON-ACUTE	19	4	23
TOTAL	66	5	71

HMO Physician Validation

No Overrides

TABLE III - 136

MEDICAL ADMISSIONS NO OVERRIDE

AEP/HMO PHYSICIAN VALIDITY

Validity Coefficients

Rule	Overall	Specific Non-Acute	Kappa
1: Lenient	81.82%	77.32%	0.56*
	(99/121)	(75/97)	
2: Moderate	69.42%	61.05%	0.39*
	(84/121)	(58/95)	
3: Conservative	38.02%	20.21%	0.08
	(46/121)	(19/94)	

* Kappa-statistic significant, $p < .01$.

TABLE III - 137

MEDICAL ADMISSIONS NO OVERRIDE

AEP/HMO VALIDITY

Lenient Non-Acute Rule

HMO PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	24	5	29
NON-ACUTE	17	75	92
TOTAL	41	80	121

TABLE III - 138

MEDICAL ADMISSIONS NO OVERRIDE

AEP/HMO VALIDITY

Moderate Non-Acute Rule

HMO PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	26	3	29
NON-ACUTE	34	58	92
TOTAL	60	61	121

TABLE III - 139

MEDICAL ADMISSIONS NO OVERRIDE

AEP/HMO VALIDITY

Conservative Non-Acute Rule

HMO PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	27	2	29
NON-ACUTE	73	19	92
TOTAL	100	21	121

TABLE III - 140

MEDICAL DAYS OF CARE NO OVERRIDE

AEP/HMO PHYSICIAN VALIDITY

Validity Coefficients

Rule	Overall	Specific Non-Acute	Kappa
1: Lenient	76.02% (428/563)	71.40% (337/472)	0.42*
2: Moderate	74.96% (422/563)	66.27% (277/418)	0.47*
3: Conservative	58.97% (332/563)	40.16% (155/386)	0.26*

* Kappa-statistic significant, $p < .01$.

TABLE III - 141

MEDICAL DAYS OF CARE NO OVERRIDE

AEP/HMO VALIDITY

Lenient Non-Acute Rule

HMO PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	91	106	197
NON-ACUTE	29	337	366
TOTAL	120	443	563

TABLE III - 142

MEDICAL DAYS OF CARE NO OVERRIDE

AEP/HMO VALIDITY

Moderate Non-Acute Rule

HMO PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	145	52	197
NON-ACUTE	89	277	366
TOTAL	234	329	563

TABLE III - 143

MEDICAL DAYS OF CARE NO OVERRIDE

AEP/HMO VALIDITY

Conservative Non-Acute Rule

HMO PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	177	20	197
NON-ACUTE	211	155	366
TOTAL	388	175	563

TABLE III - 144

DAYS WITHIN SURGICAL ADMISSIONS NO OVERRIDE

AEP/HMO PHYSICIAN VALIDITY

	Validity Coefficients		
Rule	Overall	Specific Non-Acute	Kappa
1: Lenient	67.61% (48/71)	48.89% (22/45)	0.40*
2: Moderate	77.46% (55/71)	54.29% (19/35)	0.53*
3: Conservative	77.46% (55/71)	44.83% (13/29)	0.46*

* Kappa-statistic significant, $p < .01$.

TABLE III - 145

DAYS WITHIN SURGICAL ADMISSIONS NO OVERRIDE

AEP/HMO VALIDITY

Lenient Non-Acute Rule

HMO PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	26	22	48
NON-ACUTE	1	22	23
TOTAL	27	44	71

TABLE III - 146

DAYS WITHIN SURGICAL ADMISSIONS NO OVERRIDE

AEP/HMO VALIDITY

Moderate Non-Acute Rule

HMO PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	36	12	48
NON-ACUTE	4	19	23
TOTAL	40	31	71

TABLE III - 147

DAYS WITHIN SURGICAL ADMISSIONS NO OVERRIDE

AEP/HMO VALIDITY

Conservative Non-Acute Rule

HMO PHYSICIANS' ASSESSMENT

AEP	ACUTE	NON-ACUTE	TOTAL
ACUTE	42	6	48
NON-ACUTE	10	13	23
TOTAL	52	19	71